

Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2007-13

**Multipower Variation Under Market
Microstructure Effects**

Carla Ysusi
Banco de México

October 2007

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Multipower Variation Under Market Microstructure Effects*

Carla Ysusi[†]
Banco de México

Abstract

The asymptotic theories used to estimate the integrated variance using realised variance or multipower variation suggest that returns should be sampled at the highest possible frequency. This leads to a bias problem due to market microstructure effects that can completely invalidate the theory. There is a trade-off between bias and variance when choosing the sample frequency. There is an urgent need for estimators of integrated variance that are unbiased and efficient under these effects. In this paper, multipower variation is studied under this perspective and alternative estimators are defined using the subsampling and averaging method.

Keywords: Multipower variation, Microstructure noise, Stochastic volatility models, Semimartingale, High-frequency data.

JEL Classification: C13, C51, G19.

Resumen

La teoría asintótica necesaria para estimar la varianza integrada utilizando la varianza realizada o la variación multipoder implica que los retornos deben ser muestreados a la máxima frecuencia posible. Esto conlleva a un problema de sesgo, debido a los efectos de microestructura, que pueden invalidar la teoría. Dependiendo de la frecuencia seleccionada, existen problemas de sesgo o de varianza, por lo que son necesarios estimadores que sean insesgados y eficientes bajo estos efectos. En este documento se estudia desde esta perspectiva la variación multipoder y se proponen otros estimadores basados en el método de submuestreo y promedio.

Palabras Clave: Variación multipoder, Ruido de microestructura, Modelos de volatilidad estocástica, Semimartingala, Datos en alta frecuencia.

*I am grateful to Neil Shephard for his advice and support. I wish to thank Valentina Corradi and Matthias Winkel for their helpful suggestions.

[†] Dirección General de Investigación Económica. Email: cysusi@banxico.org.mx.

1. Introduction

Realised variance and multipower variation may suffer from a bias problem due to autocorrelation in the intra-day returns. It has many sources referred to as market microstructure effects. These effects induce serial correlation in high-frequency returns, used to calculate realised variance or multipower variation; therefore they have an impact on the integrated variance estimation. To reduce this bias, small values of M (number of intra-day observations) can be chosen, but by sampling at low frequencies, we do not incorporate all the information in the data and our estimators of integrated variance will be inefficient/inconsistent. Realised variance gives a perfect estimation of integrated variance when prices are observed in continuous time, hence its calculation should be based on returns that are sampled at the highest possible frequency. So there is a trade-off between bias and variance; bias due to market frictions when sampling at high frequencies and variance due to the asymptotic assumptions that do not hold when sampling at low frequencies. The asymptotic results are based on the idea of samples of increasingly higher frequencies hence the presence of market microstructure effects can potentially invalidate them.

True prices could not equal the observed prices due to the interpolation method or market frictions. Equidistant price data must be interpolated from the observed prices and an error can arise from the econometric method used to construct this artificial price series (previous-tick or interpolation methods). Market microstructure noise can have many different origins. For stock indices the serial correlation can be caused by non-synchronous trading (Lo and MacKinlay (1990)). When individual securities are not traded simultaneously, they incorporate shocks non-synchronously to the common factor that is driving their price. This results in correlated price changes at the index level. For liquid assets the bid/ask bounce (Roll (1984)) induces negative serial correlation. When there is no new information arriving at a given moment in time, the price bounces between bid and ask prices. This effect can be strong in high-frequency data. For less liquid assets inactive trading causes a positive serial correlation. Transaction costs, misrecorded prices and the discrete nature of data that implies rounding errors may also contribute to this effect. See Andersen, Bollerslev, Diebold and Labys (1999), Andersen, Bollerslev, Diebold and Labys (2000), Bai, Russell and Tiao (2001) and Oomen (2002) for a more complete description of the bias caused by market microstructure effects.

Because market microstructure effects are present in virtually all price series, we need to find a way for these effects to have a negligible impact on the estimation of actual variance when using high-frequency data.

A first approach is the selection of an optimal sampling frequency that minimizes the bias. Bandi and Russell (2003), Hansen and Lunde (2006) and Aït-Sahalia, Mykland and Zhang (2005) study different methods to obtain such a frequency. It needs to be high enough to produce a volatility estimate with negligible sampling variation, yet low enough to avoid market microstructure bias. In Andersen, Bollerslev, Diebold and Labys (1999) the volatility signature plot was introduced to provide some initial guidance. This is a plot of average realised variance against sampling frequency; the bias is expected to increase at high-frequency levels.

A second approach is the use of alternative estimators for the integrated variance which are unbiased

in the presence of market microstructure effects. Zhang, Mykland and Aït-Sahalia (2005) developed a class of such estimators, based on realised variance and subsampling and averaging. They propose and compare different estimation methods: 1) calculating realised variance at the highest possible frequency and completely ignoring the noise, 2) sampling sparsely at a lower frequency, 3) using the optimal sampling frequency, 4) using the subsampling and averaging method, 5) bias-correcting the subsampling and averaging method. Andersen, Bollerslev, Diebold and Ebens (2001), Hansen and Lunde (2006), Oomen (2004) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2006) give other approaches to correct the bias for the realised variance. These approaches have succeeded in reducing the bias but none of them have completely solved the problem. Here, we will study alternative estimators based on realised multipower variation given that adjacent observations may reduce the bias without losing too much information. Also we will use the sampling and averaging technique with multipower variation as it may improve previous results and the estimators may turn out to be more robust to noise.

In this paper, we shall first set our Stochastic Volatility Model and describe how the market microstructure effect will be incorporated into it. The contamination due to market microstructure effects will be treated as that of an observation error. We shall then be in a position to assess realised bipower, tripower and quadpower variation in the presence of these effects. We shall thereafter introduce and study new estimators based on multipower variation and the subsampling and averaging method.

1.1. Stochastic Volatility Model and multipower variation

A standard model in financial economics is a stochastic volatility (SV) model for log-prices Y_t which follows the equation

$$Y_t = \int_0^t a_u du + \int_0^t \sigma_s dW_s, \quad t \geq 0, \quad (1)$$

where $A_t = \int_0^t a_u du$. The processes σ_t and A_t are assumed to be stochastically independent of the standard Brownian motion W . Here σ_t is called the instantaneous or spot volatility, σ_t^2 the corresponding spot variance and A_t the mean process.

More generally A_t is assumed to have locally bounded variation paths and it is set that $M_t = \int_0^t \sigma_s dW_s$, with the added condition that $\int_0^t \sigma_s^2 ds < \infty$ for all t . This is enough to guarantee that M_t is a local martingale. So the original equation (1) can be decomposed as $Y_t = A_t + M_t$. Under these assumptions Y_t is a semimartingale (see Protter(1990)). If additionally A_t is continuous then Y_t is a member of the continuous stochastic volatility semimartingale ($SVSM^c$) class.

Here, a key role is played by the integrated variance

$$\sigma_t^{2*} = \int_0^t \sigma_s^2 ds,$$

and the quadratic variation

$$[Y]_t = p \lim_{n \rightarrow \infty} \sum_{j=1}^n (Y_{t_j} - Y_{t_{j-1}})^2$$

for any sequence of partitions $t_0^{(n)} = 0 < t_1^{(n)} < \dots < t_n^{(n)} = t$ with $\sup_j \{t_j^{(n)} - t_{j-1}^{(n)}\} \rightarrow 0$ for $n \rightarrow \infty$. As A_t

is assumed to be continuous and of finite variation we obtain that

$$[Y]_t = [A]_t + 2[A, M]_t + [M]_t = [M]_t = \int_0^t \sigma_u^2 du$$

where

$$[X, Y]_t = p \lim_{n \rightarrow \infty} \sum_{j=1}^n (X_{t_j} - X_{t_{j-1}})(Y_{t_j} - Y_{t_{j-1}}).$$

This holds since the quadratic variation of any continuous, locally bounded variation process is zero (see Hull and White (1987)). If high-frequency financial data is available, the realised variance process $[Y_\delta]_t^{[2]} = \sum_{j=1}^{\lfloor t/\delta \rfloor} y_j^2$, can be defined where $y_j = Y_{j\delta} - Y_{(j-1)\delta}$ for $j = 1, 2, 3, \dots, \lfloor t/\delta \rfloor$ are the returns, given that we have observations every $\delta > 0$ periods of time.

The relationship between realised variance and quadratic variation is well known to be

$$[Y_\delta]_t^{[2]} \xrightarrow{p} [Y]_t = \int_0^t \sigma_s^2 ds$$

if $Y \in SVSM^c$.

Realised variance has been used in financial econometrics for many years, examples include Rosenberg (1972), Merton (1980), Poterba and Summers (1986), Schwert (1989), Hsieh (1991), Zhou (1996), Taylor and Xu (1997), Christensen and Prabhala (1998), Andersen, Bollerslev, Diebold and Ebens (2001), Andersen, Bollerslev, Diebold and Labys (2001). Recent literature using quadratic variation for semimartingales has been the independently and concurrently development of Andersen and Bollerslev (1998), Comte and Renault (1998) and Barndorff-Nielsen and Shephard (2001). In Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen and Shephard (2003) and Barndorff-Nielsen and Shephard (2004b) the previous theory has been extended to a Central Limit Theorem (CLT). In these papers the CLT is presented under somewhat restrictive assumptions. Recently Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2005) and Barndorff-Nielsen, Graversen, Jacod and Shephard (2006) give weaker conditions on the log-price process which ensure that the CLT holds. Many other papers on realised variance exist which are discussed in Dacorogna et al. (2001) and in the reviews by Andersen, Bollerslev and Diebold (2005) and Barndorff-Nielsen and Shephard (2007).

1.2. Modelling market microstructure noise

Empirically, it is well known that the price process is contaminated by market microstructure effects. This is why when using asymptotic results, their implementations using extremely large values of M can give misleading answers due to noise accumulation. The larger the value of M , the closer we get to the asymptotic results but also the more market microstructure effects disturb the real process.

In this paper we will try to combat the microstructure effect problem by using the skipped version of the realised bipower variation, and by taking more adjacent observations into account, i.e., realised tripower variation and realised quadpower variation. Also new estimators based on subsampling and averaging will be defined. To assess the effectiveness of these estimators in the presence of microstructure noise we need to model a contaminated process.

The additive model is a popular one, which assumes that the noise is independent and identically distributed across time and also independent of the true price process. This model has been analysed by

Bandi and Russell (2003), Corsi, Zumbach, Müller and Dacorogna (2001), Hansen and Lunde (2006) and Zhang, Mykland and Aït-Sahalia (2005). We will use it for simplicity although more general models have been already proposed by Aït-Sahalia, Mykland and Zhang (2005), Hansen and Lunde (2004) and Oomen (2002).

Let us define the contaminated log-price process as

$$\tilde{Y}_t = Y_t + \eta_t$$

where Y_t is the true log-price, η_t the microstructure noise and \tilde{Y}_t the observed log-prices. The returns are defined as

$$\begin{aligned}\tilde{Y}_{j\delta} - \tilde{Y}_{(j-1)\delta} &= \{Y_{j\delta} - Y_{(j-1)\delta}\} + \{\eta_{j\delta} - \eta_{(j-1)\delta}\} \\ \tilde{y}_j &= y_j + \varepsilon_j\end{aligned}$$

for $j = 1, 2, \dots, \lfloor t/\delta \rfloor$.

Given a fixed time period $h > 0$ (here h denotes the period of a day) with $\lfloor t/\delta \rfloor = M$ intra- h returns, let us now define

$$\tilde{y}_{j,i} = \tilde{Y}_{(i-1)h+j\delta} - \tilde{Y}_{(i-1)h+(j-1)\delta}$$

for $j = 1, 2, \dots, M$. Hence $\tilde{y}_{j,i}$ is the j -th intra-day observed return for the i -th day and can also be seen as

$$\tilde{y}_{j,i} = y_{j,i} + \varepsilon_{j,i}.$$

For the previous model we will impose the following assumptions

- (1) The random shocks η_j are Gaussian i.i.d. with mean zero and variance σ_η^2 .
- (2) $y_{j,i} \perp \varepsilon_{j,i} \quad \forall i, j$.
- (3) The true log-price process Y_t will follow the SV model (equation (1)).

Let us now simulate one thousand days of a contaminated log-price process. The true process for σ^2 will be based on the CEV process (specifically a Feller (1951) or Cox, Ingersoll and Ross (1985) square root process) where $A = 0$ and the leverage effect is ruled out.

The true returns and the contaminated returns of the first five days of our simulated series are shown in Figure 1 and 2 with $\sigma_\eta^2 = 0.0001$ and $\sigma_\eta^2 = 0.001$ respectively. In both figures we plot the returns for $M=12$, $M=72$ and $M=288$. It can easily be seen that as the value of M gets larger, the series are substantially more affected by the market microstructure noise, i.e. higher frequencies are more affected than lower frequencies. Obviously the difference between the true and the observed returns also depends on the variance of the noise. In Figure 2, where the variance of the noise is quite big, the noise completely hides the true process when $M=288$.

The problems caused by the market microstructure noise will depend on σ_η^2 hence it is important to know some empirical values. For this we can use the empirical noise-to-signal ratio, the ratio between the noise variance and the estimated average integrated variance. In Bandi and Russell (2003) this ratio equals 0.0002829 for IBM stock prices; in Hansen and Lunde (2006) it equals 0.000177 for Alcoa Inc stock prices. To be coherent with these ratios in our simulations we need σ_η^2 to be around 0.0001. Therefore we will study the cases where $\sigma_\eta^2 = 0.00001$, $\sigma_\eta^2 = 0.0001$ and $\sigma_\eta^2 = 0.001$.

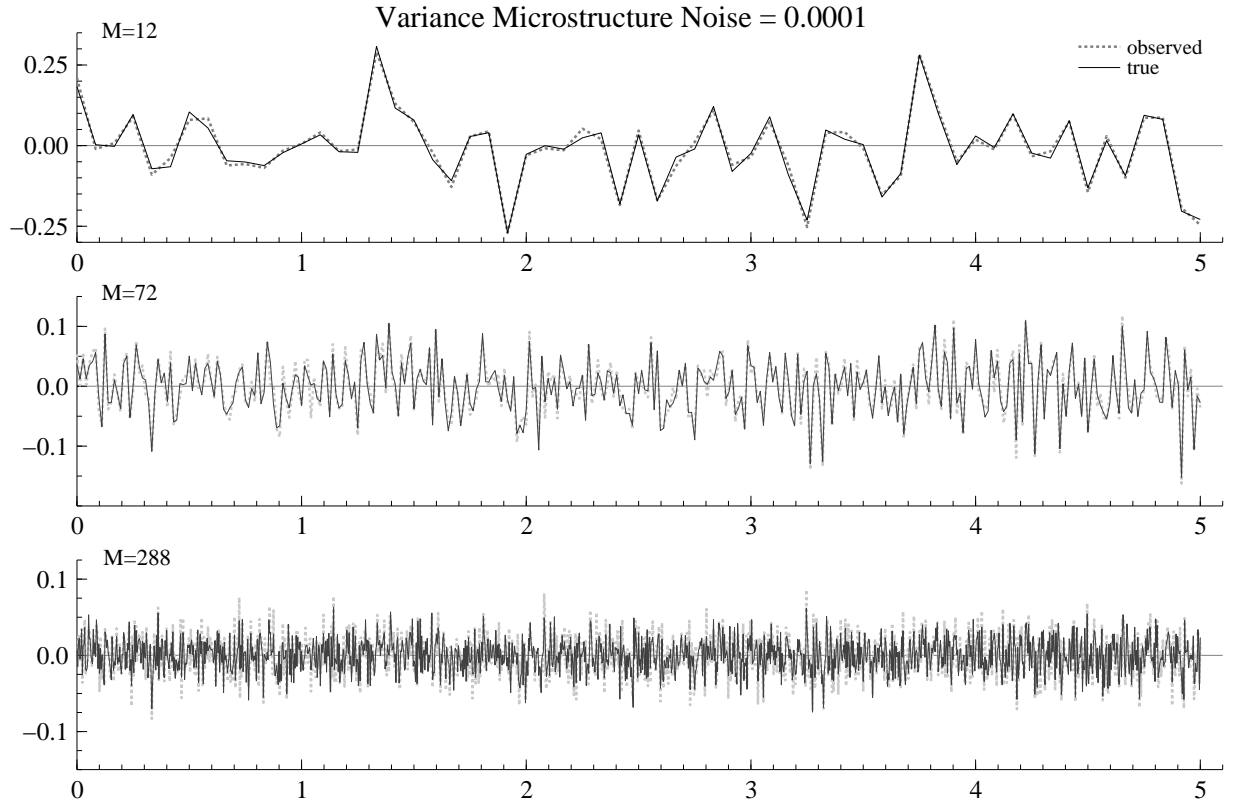


Figure 1. Plot of the true and observed return process with $\sigma_\eta^2 = 0.0001$.

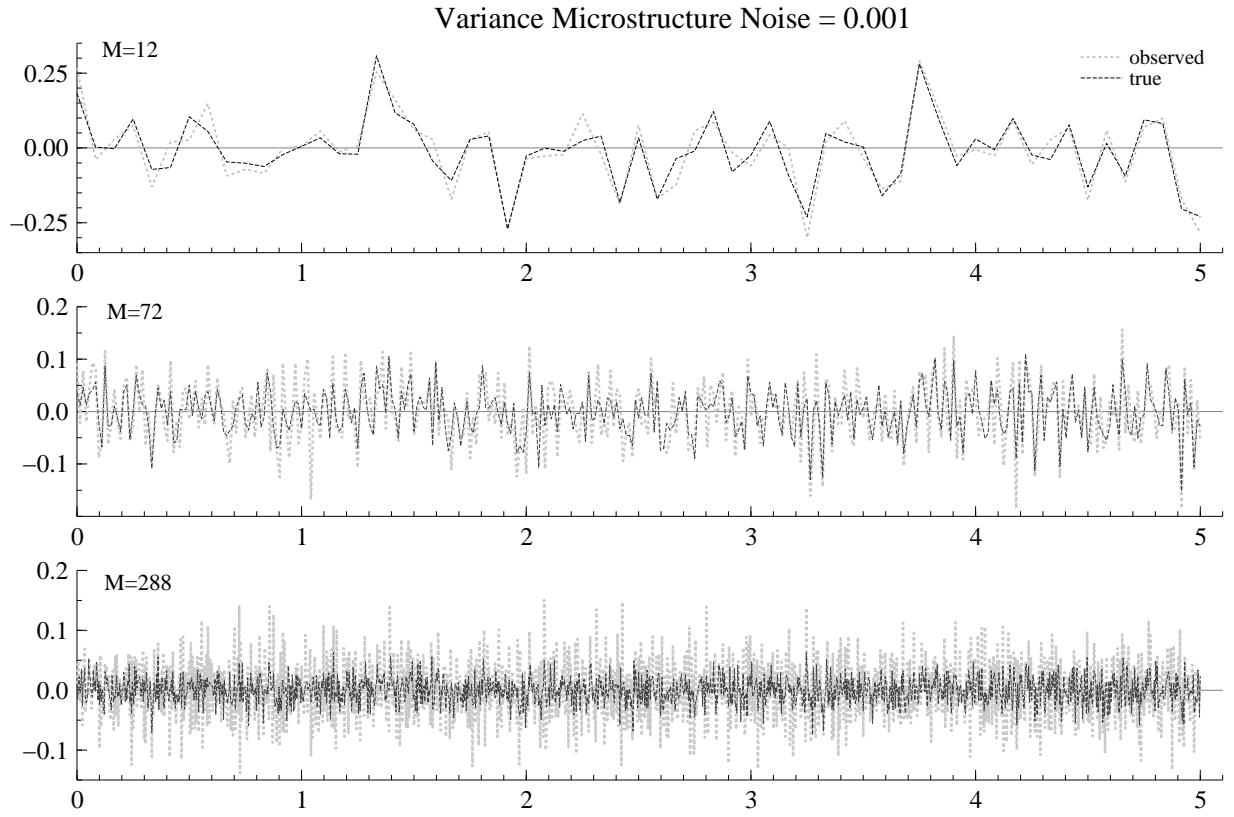


Figure 2. Plot of the true and observed return process with $\sigma_\eta^2 = 0.001$.

2. Bipower, tripower and quadpower variation in the presence of market microstructure noise

Realised bipower variation has recently been used to split the components of quadratic variation into one due to the continuous component and one due to the jump component of log-price processes. This allows us to test for the presence of jumps (see Barndorff-Nielsen and Shephard (2004a and 2006)). Realised tripower, quadpower and the skipped version of bipower variation were studied in Ysusi(2006) as alternative estimators which can be used to test for jumps in the price process.

An important issue to investigate is whether these estimators still give adequate results when market microstructure noise is present. The fact of skipping observations (skipped version of bipower variation) or using a number of adjacent observations to compute the estimators (tripower and quadpower variation) may help to reduce the bias caused by market microstructure effects. Using simulated contaminated data we shall compute realised bipower, tripower and quadpower variation and the skipped version of bipower variation and thereby attempt to assess the accuracy of these contaminated estimations.

2.1. Signature plots

As a first approach we will use signature plots where the average values of the estimators are displayed for different sampling frequencies. The sampling frequencies measured in minutes will be displayed on a logarithmic scale in our signature plots. If the estimators are affected by market microstructure effects, the bias will get bigger as the sampling frequency increases because these effects induce serial correlation in the high-frequency returns.

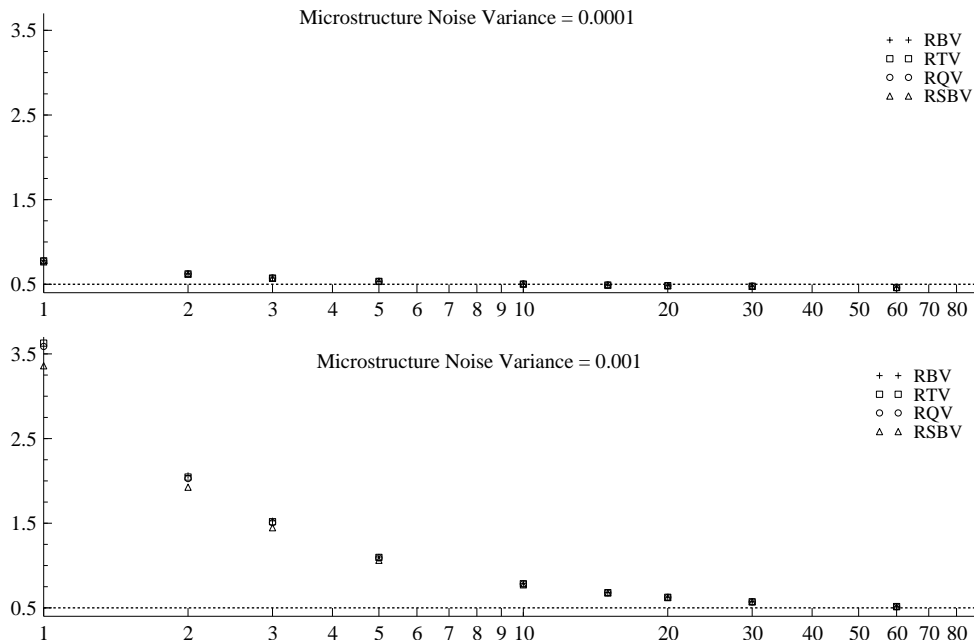


Figure 3. Signature plot when microstructure noise is present for simulated data.

This problem is evident from Figure 3. The size of the bias will depend on the variance of the noise. When the variance is big (lower graph) all the estimators are biased for sampling frequencies

above thirty minutes. There does not seem to be an evident difference between the estimators used. Only the skipped version of the realised bipower variation gives a slight improvement in the highest frequencies, but even for this estimator the bias is quite severe and it will not give reliable results.

Bias is still present in the highest frequencies (above ten minutes) when using a smaller variance for the noise (upper graph) although it seems to be quite smaller. Here there does not seem to be much difference between the estimators in any of the sampling frequencies.

It is impossible to say from the signature plots whether one estimator is more robust to market microstructure noise than the others, hence we need to study them further.

2.2. Finite sample behaviour

As the signature plots just gave us a preliminary insight into the problem of autocorrelation in our estimators when market microstructure noise is present, we will now assess the accuracy of the mixed normal asymptotic approximation to their distribution in the case of contaminated observations.

In Ysusi(2006) the finite sample behaviour of realised bipower, tripower and quadpower variation was studied when computed with true data. Here we shall focus again on their finite sample behaviour but this time in the case when they are computed with contaminated data.

If market microstructure noise did not have any effect on our estimators, we would have the following limit distribution for the realised bipower variation error

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left\{ \mu_1^{-2} \sum_{j=1}^{\lfloor t/\delta \rfloor - 1} |\tilde{y}_j| |\tilde{y}_{j+1}| - \int_0^t \sigma_s^2 ds \right\} \xrightarrow{L} N(0, \nu_{BV})$$

where $\nu_{BV} \simeq 2.60907$,

for the realised tripower variation error

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left\{ \mu_{2/3}^{-3} \sum_{j=1}^{\lfloor t/\delta \rfloor - 2} |\tilde{y}_j|^{2/3} |\tilde{y}_{j+1}|^{2/3} |\tilde{y}_{j+2}|^{2/3} - \int_0^t \sigma_s^2 ds \right\} \xrightarrow{L} N(0, \nu_{TV})$$

where $\nu_{TV} \simeq 3.0613$,

for the realised quadpower variation error

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left\{ \mu_{1/2}^{-4} \sum_{j=1}^{\lfloor t/\delta \rfloor - 3} \sqrt{|\tilde{y}_j| |\tilde{y}_{j+1}| |\tilde{y}_{j+2}| |\tilde{y}_{j+3}|} - \int_0^t \sigma_s^2 ds \right\} \xrightarrow{L} N(0, \nu_{QV})$$

where $\nu_{QV} \simeq 3.37702$

and for the skipped version of realised bipower variation error

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left\{ \mu_1^{-2} \sum_{j=1}^{\lfloor t/\delta \rfloor - 2} |\tilde{y}_j| |\tilde{y}_{j+2}| - \int_0^t \sigma_s^2 ds \right\} \xrightarrow{L} N(0, \nu_{SBV})$$

where $\nu_{SBV} \simeq 2.60907$.

M	Bs	BV SD	Cov	Bs	TV SD	Cov	Bs	QV SD	Cov	Bs	SBV SD	Cov
10^{-5}												
24	-.11	1.01	95.4	-.11	1.01	95.0	-.11	1.02	95.5	-.13	1.01	95.6
72	-.06	1.02	94.8	-.06	1.04	94.4	-.05	1.04	94.3	-.07	1.05	94.0
144	-.03	.97	96.3	-.03	.97	96.0	-.02	.98	95.5	-.01	1.01	95.0
288	.01	.98	94.6	.10	1.00	94.4	.09	1.02	95.0	.11	1.02	94.3
720	.51	1.02	91.0	.47	1.03	91.7	.45	1.03	92.4	.52	1.01	92.3
1440	1.50	1.11	67.0	1.37	1.10	70.7	1.29	1.09	72.4	1.49	1.12	67.3
10^{-4}												
24	-.08	1.00	95.1	-.09	1.01	95.2	-.09	1.01	95.7	-.09	1.01	95.3
72	.08	1.05	94.3	.08	1.05	93.7	.07	1.06	93.9	.08	1.06	93.6
144	.41	1.04	91.6	.39	1.03	91.4	.38	1.05	90.9	.43	1.06	91.0
288	1.32	1.13	72.5	1.23	1.14	74.3	1.17	1.15	77.0	1.31	1.17	71.5
720	5.39	1.89	1.7	4.95	1.81	2.8	4.69	1.77	4.6	5.22	1.80	1.5
1440	15.6	4.05	0	14.3	3.78	0	13.5	3.62	0	14.8	3.81	0
10^{-3}												
24	.19	1.08	93.1	.15	1.08	93.3	.12	1.09	93.5	.18	1.09	92.5
72	1.59	1.41	63.3	1.45	1.39	66.4	1.39	1.38	69.5	1.54	1.38	65.1
144	4.86	2.11	6.1	4.48	2.03	8.3	4.25	1.99	11.3	4.66	1.99	7.3
288	14.2	4.04	0	13.1	3.82	0	12.2	3.66	0	13.3	3.78	0
720	57.6	14.5	0	52.5	13.3	0	49.4	12.6	0	52.5	12.9	0
1440	164	38.9	0	150	35.4	0	141	33.3	0	148	34.5	0

Table 1: *Bias, standard deviation and coverage rate of the infeasible standardised realised bipower, tripower, quadpower and skipped bipower variation error in the presence of microstructure noise.*

In Table 1 the bias, standard error and coverage rate of the previous infeasible errors are recorded for $\sigma_\eta^2 = 0.00001$, $\sigma_\eta^2 = 0.0001$ and $\sigma_\eta^2 = 0.001$. From this table we can see that realised tripower and quadpower variation seem to be the least affected by the market microstructure noise. Nevertheless, when the variance of the noise is too big the noise completely hides the true process and with high sampling frequencies none of the estimators give adequate results. When the noise is not that big, there are still some problems beyond 5 minutes returns. In all the cases low sampling frequencies do not seem to be heavily affected by the noise but our theory is based on asymptotic assumptions hence low frequencies lead to inefficient estimations. So far none of our estimators seem to solve the problem of microstructure noise if the noise's variance is sufficiently big.

2.3. Test for jumps in the presence of noise

The main interest of bipower variation is that a test for jumps in the price process can be established by subtracting realised variance from realised bipower variation. It can also be performed with realised tripower or quadpower variation. When market frictions are present the real power of this test is unknown as both the realised variance and the realised bipower (tripower or quadpower) variation are affected. We shall study the effect of market microstructure noise on the test by applying it to simulated contaminated data.

If the test is not affected by market microstructure effects then the convergence theorems in Ysusi (2006) should hold. An infeasible ratio test for jumps can be based on these results using each one of

our estimators.

M	Bias	24 SD	Cove	Bias	144 SD	Cove	Bias	1440 SD	Cove
10^{-5}									
RBV	−.308	1.06	89.6	−.099	.99	93.3	.982	1.15	97.8
RTV	−.258	1.02	91.1	−.070	0.98	93.9	.663	1.16	97.8
RQV	−.256	1.03	91.3	−.073	0.99	94.3	.499	1.14	97.5
RSBV	−.327	1.06	87.9	−.084	.99	93.8	−.004	0.96	95.6
10^{-4}									
RBV	−.283	1.05	89.2	.179	1.02	95.8	3.72	1.65	99.9
RTV	−.238	1.03	91.4	.101	1.01	96.3	2.45	1.32	99.9
RQV	−.230	1.04	92.7	.055	1.02	96.3	1.84	1.21	99.7
RSBV	−.317	1.08	88.4	−.112	1.01	94.2	.014	.969	95.8
10^{-3}									
RBV	−.101	1.04	92.4	1.03	1.02	99.1	5.71	0.89	100
RTV	−.098	1.05	93.3	.627	1.04	98.7	3.69	0.98	100
RQV	−.103	1.06	93.6	.445	1.04	98.1	2.77	1.01	100
RSBV	−.292	1.09	88.8	−.152	.996	93.5	.014	0.96	96.1

Table 2: Bias, standard deviation and coverage rate of the infeasible ratio test for jumps in the presence of microstructure noise.

M	Bias	24 SD	Cove	Bias	144 SD	Cove	Bias	1440 SD	Cove
$\sigma_\eta^2 =$	10^{-4}								
50%									
RBV	−.426	1.14	85.0	−.327	1.36	85.5	3.18	1.74	99.7
RTV	−.371	1.09	87.3	−.344	1.26	85.7	1.97	1.41	99.3
RQV	−.350	1.08	87.6	−.342	1.21	86.8	1.40	1.28	98.9
RSBV	−.479	1.16	84.2	−.578	1.38	83.6	−.501	1.25	83.4
20%									
RBV	−.291	1.07	89.5	.096	1.06	94.4	3.65	1.64	100
RTV	−.237	1.04	91.5	.023	1.02	94.4	2.36	1.31	100
RQV	−.223	1.05	91.9	−.004	1.02	94.6	1.75	1.19	99.7
RSBV	−.317	1.09	87.8	−.179	1.01	92.6	−.081	.998	94.3
$\sigma_\eta^2 =$	10^{-3}								
50%									
RBV	−.211	1.08	90.5	.915	1.04	98.5	5.63	.909	100
RTV	−.212	1.09	90.5	.573	1.05	98.1	3.61	1.01	100
RQV	−.207	1.10	90.8	.413	1.06	98.2	2.69	1.03	100
RSBV	−.384	1.14	85.6	−.174	1.01	93.3	−.055	.993	94.0
20%									
RBV	−.161	1.07	91.5	.995	1.02	99.2	5.67	0.91	100
RTV	−.162	1.07	91.5	.642	1.05	98.4	3.64	1.01	100
RQV	−.160	1.09	91.6	.477	1.05	98.7	2.72	1.03	100
RSBV	−.332	1.13	86.5	−.109	.984	94.4	−.029	0.99	94.6

Table 3: Bias, standard deviation and coverage rate of the infeasible ratio test for jumps in the presence of microstructure noise when the price process includes a jump component.

Table 2 shows the bias and standard deviation of the test statistics as well as the coverage rate of the test when there are no jumps in the real process for different values of the variance of market microstructure noise. Table 3 also shows the bias, standard deviation and coverage rate of the test but this time when a jump of given size is included. It appears from these tables that the presence of jumps is underestimated when prices are contaminated by market microstructure noise. Even when jumps are big, the noise can dominate the process and completely hide them.

When market microstructure noise is affecting the price process the test for jumps can be ineffective and give wrong answers. For low frequencies jumps are difficult to detect and for high frequencies they are hidden by the market microstructure noise. In the presence of jumps, the skipped version of realised bipower variation gives slightly better results, even though the improvement is negligible. If either the noise's variance is too big or the jumps are too small then none of our estimators will be reliable. We should therefore resort to an alternative estimator.

3. Estimators based on subsampling and averaging

3.1. General estimator of actual variance

Realised variance computed from the highest possible frequency data ought to provide the best possible estimate for actual variance but market microstructure effects can invalidate asymptotic results. Although realised bipower, tripower and quadpower variation seem to prove adequate when the variance of the market microstructure noise is small, they are unreliable whenever it is large. Zhang, Mykland and Aït-Sahalia (2005) have introduced a new estimator for the actual variance based on realised variance that gives better results. They point out that although sampling at low frequency merely reduces the impact of market microstructure effects rather than corrects them for the volatility estimations, subsampling and averaging seems to be the only way to deal with them. Hence in order to benefit from the low frequency data properties they propose to select a number of subgrids of the original grid of observation times. Then they average the estimators derived from the subgrids to obtain a new estimator which is less biased than realised variance in the presence of market microstructure effects.

The necessary assumptions are, as in the case of realised variance, that the log-prices Y_t follow the stochastic volatility model described previously (equation (1)).

They suppose that the total grid of observation times $\mathcal{G} = \{t_0, \dots, t_n\}$ is partitioned in K non-overlapping subgrids $\mathcal{G}^{(k)}$ for $k = 1, \dots, K$. This can also be seen as $\mathcal{G} = \cup_{k=1}^K \mathcal{G}^{(k)}$ where $\mathcal{G}^{(k)} \cap \mathcal{G}^{(l)} = \emptyset$ when $k \neq l$.

To select the k^{th} subgrid $\mathcal{G}^{(k)}$ they start with t_{k-1} and then pick every K th sample point after that until the end of the series, T . In other words,

$$\mathcal{G}^{(k)} = (t_{k-1}, t_{k-1+K}, t_{k-1+2K}, \dots, t_{k-1+n_k K})$$

for $k = 1, \dots, K$ where n_k is the integer making $t_{k-1+n_k K}$ the last element in the corresponding subgrid.

The number of elements in the total grid is $n + 1$ whereas each subgrid has $n_k + 1$ elements. The n_k need not be the same for all k .

Afterwards they calculate the realised variance for each subgrid, i.e.

$$[Y_{K\delta}^{(k)}]_t = \sum_{t_j, t_{j+} \in \mathcal{G}^{(k)}} (Y_{t_{j+}} - Y_{t_j})^2$$

where if $t_j \in \mathcal{G}^{(k)}$ then t_{j+} denotes the following element in $\mathcal{G}^{(k)}$.

By taking the average over all the subgrids

$$[Y_\delta]_t^{(avg)} = \frac{1}{K} \sum_{k=1}^K [Y_{K\delta}^{(k)}]_t$$

they get a new estimator of the actual variance. Given a constant K , as $n \rightarrow \infty$ and so $\delta \downarrow 0$

$$[Y_\delta]_t^{(avg)} \xrightarrow{p} [Y]_t = \int_0^t \sigma_s^2 ds.$$

3.2. New estimators based on realised variance, bipower, tripower and quadpower variation

First, let us define

$$\begin{aligned} \sigma_{2\delta,j}^2 &= \int_{2\delta j}^{2\delta(j+1)} \sigma_u^2 du, \\ \lambda_{2\delta,j}^2 &= \int_{\delta(2j+1)}^{\delta(2j+3)} \sigma_u^2 du, \\ \sigma_{\delta,j}^2 &= \int_{\delta j}^{\delta(j+1)} \sigma_u^2 du. \end{aligned}$$

Following the idea of Zhang, Mykland and Ait-Sahalia (2005) and focusing, as a first approach, on the case when $K=2$, we can define for realised variance

$$[Y_\delta]_t^{(avg)} = \frac{1}{2} \left([Y_{2\delta}^{(1)}]_t + [Y_{2\delta}^{(2)}]_t \right)$$

where

$$\begin{aligned} [Y_{2\delta}^{(1)}]_t &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 1} (Y_{(2j+2)\delta} - Y_{(2j)\delta})^2 = \sum_{j=0}^{\lfloor t/2\delta \rfloor - 1} (y_{2\delta,j}^{(1)})^2 \\ &\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 1} (\sigma_{2\delta,j} \epsilon_{2\delta,j})^2 = \sum_{j=0}^{\lfloor t/2\delta \rfloor - 1} (\sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1})^2 \end{aligned}$$

given that $(y_{2\delta,j}^{(1)})^2 = (y_{2j} + y_{2j+1})^2$ and

$$\begin{aligned}
[Y_{2\delta}^{(2)}]_t &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (Y_{(2j+3)\delta} - Y_{(2j+1)\delta})^2 = \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (y_{2\delta,j}^{(2)})^2 \\
&\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\lambda_{2\delta,j} \zeta_{2\delta,j})^2 = \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} + \sigma_{\delta,2j+2} \epsilon_{\delta,2j+2})^2
\end{aligned}$$

given that $(y_{2\delta,j}^{(2)})^2 = (y_{2j+1} + y_{2j+2})^2$.

Notice that

$$\sigma_{2\delta,j}^2 = \sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2.$$

As realised bipower, tripower and quadpower variation were constructed to obtain better estimations than realised variance when market microstructure effects were present, we may expect that new estimators using the subgrid and averaging technique on bipower, tripower and quadpower variation will also be less biased than the estimator based on realised variance. Here we define these alternative estimators.

3.2.1. Bipower Variation

Let us define the new estimator based on bipower variation,

$$[Y_\delta]_t^{[1,1](avg)} = \frac{1}{2} \left([Y_{2\delta}^{(1)}]_t^{[1,1]} + [Y_{2\delta}^{(2)}]_t^{[1,1]} \right)$$

where

$$\begin{aligned}
[Y_{2\delta}^{(1)}]_t^{[1,1]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} |Y_{(2j+2)\delta} - Y_{(2j)\delta}| |Y_{(2j+4)\delta} - Y_{(2j+2)\delta}| \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} |y_{2\delta,j}^{(1)}| |y_{2\delta,j+1}^{(1)}| \\
&\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} |\sigma_{2\delta,j} \epsilon_{2\delta,j}| |\sigma_{2\delta,j+1} \epsilon_{2\delta,j+1}| \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} |\sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1}| |\sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3}|
\end{aligned}$$

and

$$\begin{aligned}
[Y_{2\delta}^{(2)}]_t^{[1,1]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} |Y_{(2j+3)\delta} - Y_{(2j+1)\delta}| |Y_{(2j+5)\delta} - Y_{(2j+3)\delta}| \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} |y_{2\delta,j}^{(2)}| |y_{2\delta,j+1}^{(2)}|
\end{aligned}$$

$$\begin{aligned}
&\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | \lambda_{2\delta,j} \zeta_{2\delta,j} | | \lambda_{2\delta,j+1} \zeta_{2\delta,j+1} | \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} + \sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} | | \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3} + \sigma_{\delta,2j+4} \epsilon_{\delta,2j+4} | .
\end{aligned}$$

In this case notice that

$$\sigma_{2\delta,j} \sigma_{2\delta,j+1} = (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/2} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/2}.$$

3.2.2. Tripower Variation

The tripower variation case can be defined as

$$[Y_\delta]_t^{[2/3,2/3,2/3](avg)} = \frac{1}{2} \left([Y_{2\delta}^{(1)}]_t^{[2/3,2/3,2/3]} + [Y_{2\delta}^{(2)}]_t^{[2/3,2/3,2/3]} \right)$$

where

$$\begin{aligned}
[Y_{2\delta}^{(1)}]_t^{[2/3,2/3,2/3]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | Y_{(2j+2)\delta} - Y_{(2j)\delta} |^{2/3} | Y_{(2j+4)\delta} - Y_{(2j+2)\delta} |^{2/3} | Y_{(2j+6)\delta} - Y_{(2j+4)\delta} |^{2/3} \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | y_{2\delta,j}^{(1)} |^{2/3} | y_{2\delta,j+1}^{(1)} |^{2/3} | y_{2\delta,j+2}^{(1)} |^{2/3} \\
&\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | \sigma_{2\delta,j} \epsilon_{2\delta,j} |^{2/3} | \sigma_{2\delta,j+1} \epsilon_{2\delta,j+1} |^{2/3} | \sigma_{2\delta,j+2} \epsilon_{2\delta,j+2} |^{2/3} \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} | \sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} |^{2/3} | \sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3} |^{2/3} \\
&\quad | \sigma_{\delta,2j+4} \epsilon_{\delta,2j+4} + \sigma_{\delta,2j+5} \epsilon_{\delta,2j+5} |^{2/3}
\end{aligned}$$

and

$$\begin{aligned}
[Y_{2\delta}^{(2)}]_t^{[2/3,2/3,2/3]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | Y_{(2j+3)\delta} - Y_{(2j+1)\delta} |^{2/3} | Y_{(2j+5)\delta} - Y_{(2j+3)\delta} |^{2/3} | Y_{(2j+7)\delta} - Y_{(2j+5)\delta} |^{2/3} \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | y_{2\delta,j}^{(2)} |^{2/3} | y_{2\delta,j+1}^{(2)} |^{2/3} | y_{2\delta,j+2}^{(2)} |^{2/3} \\
&\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | \lambda_{2\delta,j} \zeta_{2\delta,j} |^{2/3} | \lambda_{2\delta,j+1} \zeta_{2\delta,j+1} |^{2/3} | \lambda_{2\delta,j+2} \zeta_{2\delta,j+2} |^{2/3} \\
&= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} + \sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} |^{2/3} | \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3} + \sigma_{\delta,2j+4} \epsilon_{\delta,2j+4} |^{2/3} \\
&\quad | \sigma_{\delta,2j+5} \epsilon_{\delta,2j+5} + \sigma_{\delta,2j+6} \epsilon_{\delta,2j+6} |^{2/3} .
\end{aligned}$$

Here notice that

$$\sigma_{2\delta,j}^{2/3} \sigma_{2\delta,j+1}^{2/3} \sigma_{2\delta,j+2}^{2/3} = (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/3} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/3} (\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)^{1/3}.$$

3.2.3. Quadpower Variation

Finally, let us define the following estimator based on quadpower variation,

$$[Y_\delta]_t^{[1/2,1/2,1/2,1/2](avg)} = \frac{1}{2} \left([Y_{2\delta}^{(1)}]_t^{[1/2,1/2,1/2,1/2]} + [Y_{2\delta}^{(2)}]_t^{[1/2,1/2,1/2,1/2]} \right)$$

where

$$\begin{aligned} [Y_{2\delta}^{(1)}]_t^{[1/2,1/2,1/2,1/2]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} |Y_{(2j+2)\delta} - Y_{(2j)\delta}|^{1/2} |Y_{(2j+4)\delta} - Y_{(2j+2)\delta}|^{1/2} \\ &\quad |Y_{(2j+6)\delta} - Y_{(2j+4)\delta}|^{1/2} |Y_{(2j+8)\delta} - Y_{(2j+6)\delta}|^{1/2} \\ &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} |y_{2\delta,j}^{(1)}|^{1/2} |y_{2\delta,j+1}^{(1)}|^{1/2} |y_{2\delta,j+2}^{(1)}|^{1/2} |y_{2\delta,j+3}^{(1)}|^{1/2} \\ &\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} |\sigma_{2\delta,j}\epsilon_{2\delta,j}|^{1/2} |\sigma_{2\delta,j+1}\epsilon_{2\delta,j+1}|^{1/2} |\sigma_{2\delta,j+2}\epsilon_{2\delta,j+2}|^{1/2} |\sigma_{2\delta,j+3}\epsilon_{2\delta,j+3}|^{1/2} \\ &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} |\sigma_{\delta,2j}\epsilon_{\delta,2j} + \sigma_{\delta,2j+1}\epsilon_{\delta,2j+1}|^{1/2} |\sigma_{\delta,2j+2}\epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3}\epsilon_{\delta,2j+3}|^{1/2} \\ &\quad |\sigma_{\delta,2j+4}\epsilon_{\delta,2j+4} + \sigma_{\delta,2j+5}\epsilon_{\delta,2j+5}|^{1/2} |\sigma_{\delta,2j+6}\epsilon_{\delta,2j+6} + \sigma_{\delta,2j+7}\epsilon_{\delta,2j+7}|^{1/2} \end{aligned}$$

and

$$\begin{aligned} [Y_{2\delta}^{(2)}]_t^{[1/2,1/2,1/2,1/2]} &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} |Y_{(2j+3)\delta} - Y_{(2j+1)\delta}|^{1/2} |Y_{(2j+5)\delta} - Y_{(2j+3)\delta}|^{1/2} \\ &\quad |Y_{(2j+7)\delta} - Y_{(2j+5)\delta}|^{1/2} |Y_{(2j+9)\delta} - Y_{(2j+7)\delta}|^{1/2} \\ &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} |y_{2\delta,j}^{(2)}|^{1/2} |y_{2\delta,j+1}^{(2)}|^{1/2} |y_{2\delta,j+2}^{(2)}|^{1/2} |y_{2\delta,j+3}^{(2)}|^{1/2} \\ &\stackrel{law}{=} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} |\lambda_{2\delta,j}\zeta_{2\delta,j}|^{1/2} |\lambda_{2\delta,j+1}\zeta_{2\delta,j+1}|^{1/2} |\lambda_{2\delta,j+2}\zeta_{2\delta,j+2}|^{1/2} |\lambda_{2\delta,j+3}\zeta_{2\delta,j+3}|^{1/2} \\ &= \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} |\sigma_{\delta,2j+1}\epsilon_{\delta,2j+1} + \sigma_{\delta,2j+2}\epsilon_{\delta,2j+2}|^{1/2} |\sigma_{\delta,2j+3}\epsilon_{\delta,2j+3} + \sigma_{\delta,2j+4}\epsilon_{\delta,2j+4}|^{1/2} \\ &\quad |\sigma_{\delta,2j+5}\epsilon_{\delta,2j+5} + \sigma_{\delta,2j+6}\epsilon_{\delta,2j+6}|^{1/2} |\sigma_{\delta,2j+7}\epsilon_{\delta,2j+7} + \sigma_{\delta,2j+8}\epsilon_{\delta,2j+8}|^{1/2}. \end{aligned}$$

Here we have that

$$\sigma_{2\delta,j}^{1/2} \sigma_{2\delta,j+1}^{1/2} \sigma_{2\delta,j+2}^{1/2} \sigma_{2\delta,j+3}^{1/2} = (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/4} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/4} (\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)^{1/4} (\sigma_{\delta,2j+6}^2 + \sigma_{\delta,2j+7}^2)^{1/4}.$$

3.3. Daily time series

In order to produce daily time series, we need to consider a fixed time period h (corresponding to the period of one day) with $\lfloor t/\delta \rfloor = M$ intra-day observed (and contaminated) returns, during each day, defined as

$$\tilde{y}_{j,i} = \tilde{Y}_{(i-1)h+(j+1)\delta} - \tilde{Y}_{(i-1)h+j\delta},$$

for the $j - th$ intra-day return in the $i - th$ period.

Let us point out that a bias is introduced by using finite values of M because each estimator will have a different number of components in the summation. To avoid this the following modified estimators will be used

$$\begin{aligned} [Y_M]_i^{[1,1](avg)} &= \frac{1}{2} \left(\frac{M/2}{M/2-1} [Y_{M/2}^{(1)}]_i^{[1,1]} + \frac{M/2-1}{M/2-2} [Y_{M/2}^{(2)}]_i^{[1,1]} \right), \\ [Y_M]_i^{[2/3,2/3,2/3](avg)} &= \frac{1}{2} \left(\frac{M/2}{M/2-2} [Y_{M/2}^{(1)}]_i^{[2/3,2/3,2/3]} + \frac{M/2-1}{M/2-3} [Y_{M/2}^{(2)}]_i^{[2/3,2/3,2/3]} \right), \\ [Y_M]_i^{[1/2,1/2,1/2,1/2](avg)} &= \frac{1}{2} \left(\frac{M/2}{M/2-3} [Y_{M/2}^{(1)}]_i^{[1/2,1/2,1/2,1/2]} + \frac{M/2-1}{M/2-4} [Y_{M/2}^{(2)}]_i^{[1/2,1/2,1/2,1/2]} \right), \\ [Y_M]_i^{[1,0,1](avg)} &= \frac{1}{2} \left(\frac{M/2}{M/2-2} [Y_{M/2}^{(1)}]_i^{[1,0,1]} + \frac{M/2-1}{M/2-3} [Y_{M/2}^{(2)}]_i^{[1,0,1]} \right). \end{aligned}$$

3.4. Asymptotic distributions

To assess the behaviour of our new estimators under market microstructure noise we need them to converge in distribution.

Given the assumptions in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2005) and Barndorff-Nielsen, Graversen, Jacod and Shephard (2006), we can obtain the asymptotic distributions below by setting $A = 0$.

Result 1

If $Y \in SVSM^c$ then as $\delta \downarrow 0$

$$\sqrt{\frac{1}{2\delta}} \begin{pmatrix} [Y_{2\delta}^{(1)}]_t - \int_0^t \sigma_s^2 ds \\ [Y_{2\delta}^{(2)}]_t - \int_0^t \sigma_s^2 ds \end{pmatrix} \xrightarrow{L} MN \left(0, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \int_0^t \sigma_s^4 ds \right).$$

Hence we get

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left([Y_\delta]_t^{(avg)} - \int_0^t \sigma_s^2 ds \right) \xrightarrow{L} N(0, 3).$$

Result 2

If $Y \in SVSM^c$ then as $\delta \downarrow 0$

$$\begin{aligned} \sqrt{\frac{1}{2\delta}} \left(\begin{bmatrix} [Y_{2\delta}^{(1)}]_t^{[1,1]} - \mu_1^2 \int_0^t \sigma_s^2 ds \\ [Y_{2\delta}^{(2)}]_t^{[1,1]} - \mu_1^2 \int_0^t \sigma_s^2 ds \end{bmatrix} \right) &\xrightarrow{L} MN \left(0, \begin{pmatrix} k_{1,2} & k_{2,2} \\ k_{2,2} & k_{1,2} \end{pmatrix} \int_0^t \sigma_s^4 ds \right) \\ &\simeq MN \left(0, \begin{pmatrix} 2.592 & 0.988 \\ 0.988 & 2.592 \end{pmatrix} \int_0^t \sigma_s^4 ds \right). \end{aligned}$$

Hence we get

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left([Y_\delta]_t^{[1,1](avg)} - \mu_1^2 \int_0^t \sigma_s^2 ds \right) \xrightarrow{L} N \left(0, \vartheta_{BVS} \right)$$

where $\vartheta_{BVS} = (2k_{1,2} + 2k_{2,2})/2 \simeq 3.581$.

Result 3

If $Y \in SVSM^c$ then as $\delta \downarrow 0$

$$\begin{aligned} \sqrt{\frac{1}{2\delta}} \left(\begin{bmatrix} [Y_{2\delta}^{(1)}]_t^{[2/3,2/3,2/3]} - \mu_{2/3}^3 \int_0^t \sigma_s^2 ds \\ [Y_{2\delta}^{(2)}]_t^{[2/3,2/3,2/3]} - \mu_{2/3}^3 \int_0^t \sigma_s^2 ds \end{bmatrix} \right) &\xrightarrow{L} MN \left(0, \begin{pmatrix} k_{1,3} & k_{2,3} \\ k_{2,3} & k_{1,3} \end{pmatrix} \int_0^t \sigma_s^4 ds \right) \\ &\simeq MN \left(0, \begin{pmatrix} 3.049 & 1.018 \\ 1.019 & 3.049 \end{pmatrix} \int_0^t \sigma_s^4 ds \right). \end{aligned}$$

Hence we get

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left([Y_\delta]_t^{[2/3,2/3,2/3](avg)} - \mu_{2/3}^3 \int_0^t \sigma_s^2 ds \right) \xrightarrow{L} N \left(0, \vartheta_{BTS} \right)$$

where $\vartheta_{BTS} = (2k_{1,3} + 2k_{2,3})/2 \simeq 4.067$.

Result 4

If $Y \in SVSM^c$ then as $\delta \downarrow 0$

$$\begin{aligned} \sqrt{\frac{1}{2\delta}} \left(\begin{bmatrix} [Y_{2\delta}^{(1)}]_t^{[1/2,1/2,1/2,1/2]} - \mu_{1/2}^4 \int_0^t \sigma_s^2 ds \\ [Y_{2\delta}^{(2)}]_t^{[1/2,1/2,1/2,1/2]} - \mu_{1/2}^4 \int_0^t \sigma_s^2 ds \end{bmatrix} \right) &\xrightarrow{L} MN \left(0, \begin{pmatrix} k_{1,4} & k_{2,4} \\ k_{2,4} & k_{1,4} \end{pmatrix} \int_0^t \sigma_s^4 ds \right) \\ &\simeq MN \left(0, \begin{pmatrix} 3.377 & 1.013 \\ 1.013 & 3.377 \end{pmatrix} \int_0^t \sigma_s^4 ds \right). \end{aligned}$$

Hence we get

$$\frac{1}{\delta^{1/2} \sqrt{\int_0^t \sigma_s^4 ds}} \left([Y_\delta]_t^{[1/2, 1/2, 1/2, 1/2](avg)} - \mu_{1/2}^4 \int_0^t \sigma_s^2 ds \right) \xrightarrow{L} N\left(0, \vartheta_{BQS}\right)$$

where $\vartheta_{BQS} = (2k_{1,4} + 2k_{2,4})/2 \simeq 4.389$.

The derivations of these results are given in the appendix.

3.5. Signature plots

Now, in order to assess the accuracy of our estimators we will use the simulated series. Firstly we shall check whether the use of subgridding and averaging in the calculations of the estimators reduces the bias caused by market microstructure effects.

Signature plots in logarithmic scale (Figure 4) show that the bias is still a problem when very high frequencies are used to calculate the estimators. Nevertheless, it appears that the use of subgrids and averaging reduces the bias. By comparing the upper graph in Figure 4 to the upper graph in Figure 3, we can see that now all the estimators seem robust to microstructure noise when using five minutes returns. There is still some bias for frequencies larger than two minutes but not as much as with the standard estimators. From the lower graph, where the variance of the microstructure noise is large, notice that the bias is still quite important for high frequencies but this time just for sampling frequencies above fifteen minutes. As in Figure 3, here again there is no obvious difference between the estimators. Although subgridding and averaging lowers the bias due to autocorrelation in high-frequency data, it increases the bias due to discretization in low frequency data. With ninety minutes returns a bias can be observed, certainly because of the subsampling.

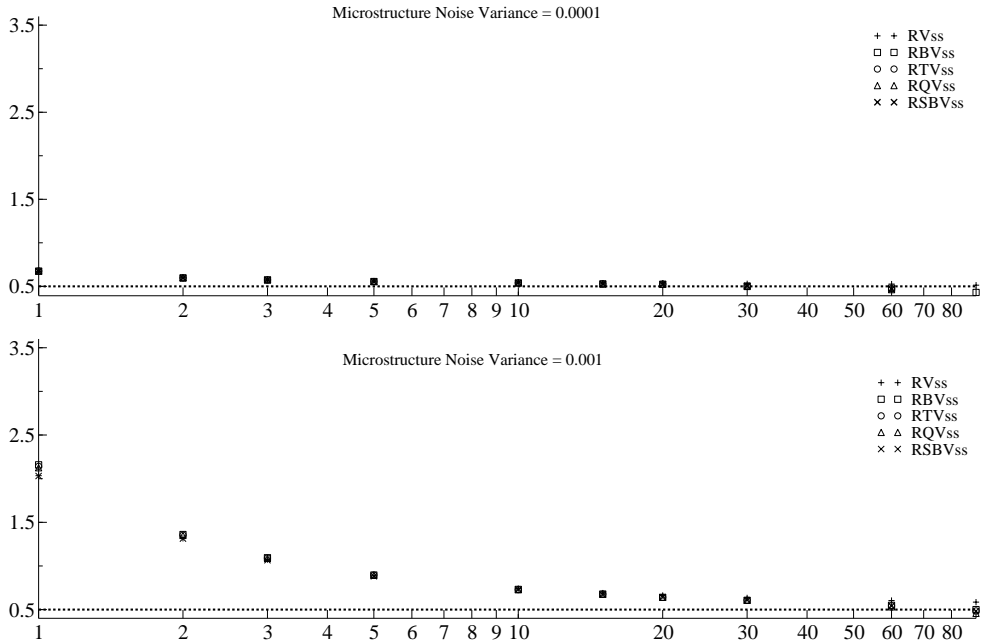


Figure 4. Signature plot for the subgridding and averaging estimators when microstructure noise is present.

3.6. Finite sample behaviour

To reinforce the results, Tables 4, 5 and 6 give an alternative and more complete view of the analysis. By using the asymptotic distributions of the estimators calculated using subgridding and averaging (Results 1, 2, 3 and 4) we can compare these new estimators with the standard ones more precisely. These tables report the bias, standard deviation and coverage rate of the infeasible standardised t-statistics.

As revealed by the asymptotic distribution, our new estimators are less efficient than the standard ones. The tables confirm this fact. When there is no market microstructure noise, the standard estimators give better results than the new estimators, i.e., they are less biased and their standard deviation is closer to one. When microstructure noise is added to the price series, the new estimators seem to be more robust for higher frequencies. When the variance of the noise equals 0.0001, the bias, standard deviation and coverage rate of the new estimators exhibit a significant improvement on frequencies above $M = 144$. Nevertheless for $M = 720$, even if the new estimators behave better than the standard ones, the bias is too big. In the case when the variance of the noise is 0.001 the new estimators will be preferred for values of M higher than 24, but the noise will completely hide the process for frequencies higher than $M = 144$. Irrespectively of the size of the microstructure noise, the estimators based on realised tripower variation and realised quadpower variation give the best results although the differences between all the estimators are very subtle.

M	Bias	RV SD	Cov	Bias	RV _{ss} SD	Cov	Bias	RBV SD	Cov	Bias	RBV _{ss} SD	Cov
0												
12	-.018	1.01	95.3	-.155	.995	96.4	-.172	.999	96.8	-.422	.953	97.2
72	-.047	1.03	94.5	-.082	.995	95.4	-.081	1.02	94.9	-.109	1.03	93.9
144	-.033	1.00	95.2	-.081	1.01	95.7	-.080	.962	96.4	-.086	1.02	94.5
288	-.013	.984	95.5	-.043	.999	96.2	-.027	.976	95.5	-.067	.996	95.8
720	-.004	.955	95.7	-.014	.982	95.7	-.015	.983	94.8	-.040	.974	95.5
1440	.041	.988	95.3	0.017	.960	96.1	.008	.978	95.3	.013	.981	95.6
10 ⁻⁴												
12	.040	1.04	95.1	-.147	.999	96.3	-.150	1.01	96.3	-.418	.952	97.3
72	.141	1.06	93.6	-.007	1.00	95.2	.082	1.05	94.3	-.046	1.03	93.6
144	.508	1.07	90.6	.145	1.03	94.6	.406	1.04	91.6	.123	1.05	94.2
288	1.53	1.15	63.5	.580	1.04	89.2	1.33	1.13	72.5	.512	1.05	90.1
720	5.99	1.91	0.2	2.42	1.21	35.8	5.39	1.88	1.7	2.23	1.21	42.3
1440	17.0	4.23	0	6.94	1.98	0	15.6	4.05	0	6.55	1.96	0.1
10 ⁻³												
12	.176	1.09	93.4	-.097	1.02	95.5	-.031	1.07	95.0	-.386	.966	97.1
72	1.85	1.41	57.1	.683	1.12	86.1	1.59	1.41	63.3	.571	1.14	88.5
144	5.36	2.11	2.8	2.13	1.34	47.1.4	4.85	2.10	6.1	1.98	1.36	52.4
288	15.3	4.11	0	6.16	2.05	0.6	14.2	4.04	0	5.87	2.11	1.5
720	60.1	14.6	0	24.4	6.05	0	57.5	14.5	0	24.1	6.22	0
1440	169	39.4	0	69.3	16.2	0	164	38.9	0	69.4	16.8	0

Table 4: Bias, standard deviation and coverage rate of the infeasible standardised realised variance error, the infeasible standardised realised bipower variation error and the subsampled ones in the absence and presence of microstructure noise.

M	Bs	RTV SD	Cov	Bs	RTV _{ss} SD	Cov	Bs	RQV SD	Cov	Bs	RQV _{ss} SD	Cov
0												
12	-.194	.998	96.4	-.518	.877	97.7	-.217	.997	96.9	-.707	.706	99.2
72	-.072	1.03	94.2	-.095	1.04	94.3	-.069	1.04	94.6	-.089	1.05	93.9
144	-.075	.963	96.2	-.073	1.05	94.5	-.070	.974	95.8	-.072	1.05	94.8
288	-.024	.990	95.1	-.057	1.00	95.3	-.026	1.01	94.6	-.052	1.02	94.7
720	-.003	.991	94.7	-.032	.964	95.9	.006	.993	94.9	-.023	.960	96.1
1440	.010	.982	95.5	.022	.980	95.8	.006	.980	95.5	.028	.985	95.3
10 ⁻⁴												
12	-.169	1.02	96.4	-.513	.875	97.7	-.193	1.02	96.4	-.699	.705	99.3
72	.078	1.06	93.7	-.037	1.04	94.2	.074	1.06	93.9	-.033	1.04	94.1
144	.391	1.04	91.4	.122	1.07	93.6	.384	1.05	90.9	.118	1.08	93.5
288	1.23	1.14	74.3	.497	1.05	90.8	1.17	1.15	77.0	.489	1.07	89.9
720	4.96	1.81	2.8	2.09	1.19	46.7	4.69	1.77	4.6	2.02	1.19	49.0
1440	14.3	3.78	0	6.13	1.89	0.1	13.55	3.62	0	5.87	1.83	0.4
10 ⁻³												
12	-.053	1.08	94.7	-.486	.889	97.6	-.080	1.09	95.1	-.682	.710	99.0
72	1.45	1.39	66.4	.524	1.15	89.1	1.39	1.38	69.5	.497	1.14	88.7
144	4.48	2.03	8.3	1.85	1.37	56.5	4.25	1.99	11.3	1.77	1.37	58.3
288	13.2	3.82	0	5.49	2.06	2.3	12.2	3.66	0	5.25	2.03	3.4
720	52.5	13.3	0	22.4	5.86	0	49.4	12.6	0	21.3	5.64	0
1440	150	35.5	0	64.3	15.6	0	141	33.3	0	61.3	14.9	0

Table 5: Bias, standard deviation and coverage rate of the infeasible standardised realised tripower variation error, the infeasible standardised realised quadpower variation error and the subsampled ones in the absence and presence of microstructure noise.

M	Bs	RSBV SD	Cov	Bs	RSBV _{ss} SD	Cov
0						
12	-.230	.968	97.1	-.551	.848	98.5
72	-.093	1.05	94.3	-.113	1.04	94.6
144	-.068	.994	94.9	-.095	1.07	94.3
288	-.023	1.01	94.6	-.057	1.03	93.5
720	.001	.970	95.2	-.010	.974	95.2
1440	.022	.991	95.2	.022	.975	95.2
10 ⁻⁴						
12	-.208	.981	97.0	-.546	.848	98.3
72	.075	1.06	93.6	-.048	1.05	94.7
144	.429	1.06	91.0	.115	1.09	92.9
288	1.31	1.68	71.5	.526	1.07	89.7
720	5.22	1.80	1.5	2.22	1.21	43.0
1440	14.8	3.81	0	6.34	1.88	0.1
10 ⁻³						
12	-.095	1.05	96.1	-.512	.863	98.1
72	1.54	1.38	65.1	.552	1.15	88.5
144	4.66	1.99	7.3	1.92	1.38	54.2
288	13.3	3.78	0	5.61	1.99	1.5
720	52.5	12.9	0	22.3	5.68	0
1440	148	34.5	0	63.3	15.0	0

Table 6: Bias, standard deviation and coverage rate of the infeasible standardised skipped version of realised bipower variation error and the subsampled ones in the absence and presence of microstructure noise.

4. Conclusions

The aim of this paper was to study market microstructure effects and determine how they affect the estimation of the actual variance when using high-frequency data. It is well known that estimators of actual variance become biased in the presence of market frictions when increasing the sampling frequency, nevertheless their asymptotic properties oblige us to sample at the highest possible frequency. Alternative approaches are needed to overcome the bias-variance trade-off.

When high-frequency data is available, the usual financial practice is to sample sparsely to reduce the bias caused by the market microstructure effect in the realised variance. Here we claimed that using adjacent observations, i.e. realised bipower, tripower and quadpower variation reduces this bias without throwing away too much information. Although these estimators improved the results given by realised variance, the market microstructure noise can completely hide the true process if its variance is big enough. Therefore we constructed alternative estimators based on the sampling and averaging technique introduced by Zhang, Mykland and Aït-Sahalia (2005). These estimators are considerably more robust to the noise although a bias still exists when using very high frequencies.

Recently research has been focused on market microstructure noise, its effects, its quantification and correction. Here we just gave some alternative estimators that improved the use of realised variance, although they did not turn out to be completely robust to the noise. Notice the important improvements achieved when using two subgrids to define the estimators. It would be interesting to determine how much can be gained by increasing the number of subgrids.

We should point out that we assumed the noise to be independent and identically distributed across time and independent of the true price process. A similar research could be carried out under a more general specification where the noise may be autocorrelated and need not be independent of the latent price process.

At the moment many efforts are concentrated on the correct estimation of integrated variance using high-frequency data. So far all the various difficulties encountered in this estimation have been addressed on the basis of realised variance. It has been the main focus of ongoing research. Nevertheless, as shown here, other estimators based on multipower variation could be as effective as realised variance, but with extra advantages such as robustness to jumps and to market microstructure effects.

5. Appendix: Derivations

Derivation of the asymptotic distributions for the subsampling and averaged based estimators

For the derivation of the asymptotic distributions of this chapter we will set the assumptions in Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2005) and Barndorff-Nielsen, Graversen, Jacod and Shephard (2006) to hold. Also we set $A = 0$.

Derivation for Realised Variance

We need to find the covariance matrix of

$$\sqrt{\frac{1}{2\delta}} \begin{pmatrix} [Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \\ [Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \end{pmatrix}.$$

Notice the following results when $\delta \downarrow 0$,

$$\begin{aligned} & \sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \\ & \stackrel{law}{=} \sqrt{\frac{1}{2\delta}} \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1})^2 - (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2) \right) \\ & = \sqrt{\frac{1}{2\delta}} \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} ((\sigma_{\delta,2j}^2 (\epsilon_{\delta,2j}^2 - 1) + \sigma_{\delta,2j+1}^2 (\epsilon_{\delta,2j+1}^2 - 1) + 2\sigma_{\delta,2j} \sigma_{\delta,2j+1} \epsilon_{\delta,2j} \epsilon_{\delta,2j+1})) \right) \\ \\ & \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} \sigma_{2\delta,j}^4 = \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^2 \\ & = \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4 + 2\sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2) \\ & \xrightarrow{p} \int_0^t \sigma_s^4 ds. \end{aligned}$$

This comes from the fact that

$$\begin{aligned} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4 + 2\sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2) &= \frac{1}{2} \left(\frac{1}{\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4) \right) + \frac{1}{2} \left(\frac{1}{\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} 2\sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2 \right) \\ &\xrightarrow{p} 1/2 \int_0^t \sigma_s^4 ds + 1/2 \int_0^t \sigma_s^4 ds \end{aligned}$$

as

$$\begin{aligned} \frac{1}{\delta} \sum_{j=0}^{\lfloor t/\delta \rfloor - 1} \sigma_{\delta,j}^4 &= \frac{1}{\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} (\sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4) \\ &\xrightarrow{p} \int_0^t \sigma_s^4 ds \end{aligned}$$

and

$$\frac{1}{\delta} \sum_{j=0}^{\lfloor t/\delta \rfloor - 2} \sigma_{\delta,j}^2 \sigma_{\delta,j+1}^2 = \frac{1}{\delta} \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} \sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2 + \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} \sigma_{\delta,2j+1}^2 \sigma_{\delta,2j+2}^2 \right)$$

$$\xrightarrow{p} \int_0^t \sigma_s^4 ds.$$

With the previous results it is easy to obtain the variance,

$$\begin{aligned} Var & \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) \\ & \stackrel{law}{=} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} \sigma_{\delta,2j}^4 Var(\epsilon_{\delta,2j}^2 - 1) + \sigma_{\delta,2j+1}^4 Var(\epsilon_{\delta,2j+1}^2 - 1) + 4\sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2 Var(\epsilon_{\delta,2j} \epsilon_{\delta,2j+1}) \\ & = 2 \left(\frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 2} \sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4 + 2\sigma_{\delta,2j}^2 \sigma_{\delta,2j+1}^2 \right) \\ & \xrightarrow{p} 2 \int_0^t \sigma_s^4 ds. \end{aligned}$$

Analogously, we can obtain that

$$Var \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) \xrightarrow{p} 2 \int_0^t \sigma_s^4 ds.$$

For the covariance, we have that

$$\begin{aligned} Cov & \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right), \sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) \\ & = E \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \cdot \sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) - \\ & \quad E \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) E \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) \\ & \stackrel{law}{=} \frac{1}{2\delta} E \left(\left\{ \left(\sigma_{\delta,0}^2 (\epsilon_{\delta,0}^2 - 1) + \sigma_{\delta,1}^2 (\epsilon_{\delta,1}^2 - 1) + 2\sigma_{\delta,0} \sigma_{\delta,1} \epsilon_{\delta,0} \epsilon_{\delta,1} \right) + \left(\sigma_{\delta,2}^2 (\epsilon_{\delta,2}^2 - 1) + \sigma_{\delta,3}^2 (\epsilon_{\delta,3}^2 - 1) \right. \right. \right. \\ & \quad \left. \left. \left. + 2\sigma_{\delta,2} \sigma_{\delta,3} \epsilon_{\delta,2} \epsilon_{\delta,3} \right) + \dots \right\} \cdot \left\{ \left(\sigma_{\delta,1}^2 (\epsilon_{\delta,1}^2 - 1) + \sigma_{\delta,2}^2 (\epsilon_{\delta,2}^2 - 1) + 2\sigma_{\delta,1} \sigma_{\delta,2} \epsilon_{\delta,1} \epsilon_{\delta,2} \right) \right. \right. \\ & \quad \left. \left. \left. + \left(\sigma_{\delta,3}^2 (\epsilon_{\delta,3}^2 - 1) + \sigma_{\delta,4}^2 (\epsilon_{\delta,4}^2 - 1) + 2\sigma_{\delta,3} \sigma_{\delta,4} \epsilon_{\delta,3} \epsilon_{\delta,4} \right) + \dots \right\} \right) \\ & = \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} \sigma_{\delta,2j}^4 E(\epsilon_{\delta,2j}^2 - 1)^2 + \sigma_{\delta,2j+1}^4 E(\epsilon_{\delta,2j+1}^2 - 1)^2 \\ & = 2 \frac{1}{2\delta} \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 3} \sigma_{\delta,2j}^4 + \sigma_{\delta,2j+1}^4 \right) \\ & \xrightarrow{p} \int_0^t \sigma_s^4 ds \end{aligned}$$

Then

$$Cov \left(\sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(1)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right), \sqrt{\frac{1}{2\delta}} \left([Y_{2\delta}^{(2)}]_t^{[2]} - \int_0^t \sigma_s^2 ds \right) \right) \xrightarrow{p} \int_0^t \sigma_s^4 ds$$

Derivation for Realised Bipower Variation

We need to find the covariance matrix of

$$\sqrt{\frac{1}{2\delta}} \left(\mu_1^{-2} [Y_{2\delta}^{(1)}]_t^{[1,1]} - \int_0^t \sigma_s^2 ds \right)$$

If $\delta \downarrow 0$, we have that

$$\begin{aligned} Var \left(\sqrt{\frac{1}{2\delta}} \left(\mu_1^{-2} [Y_{2\delta}^{(1)}]_t^{[1,1]} - \int_0^t \sigma_s^2 ds \right) \right) \\ \stackrel{law}{=} \frac{1}{2\delta} \mu_1^{-4} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | y_{2\delta,j}^{(1)} || y_{2\delta,j+1}^{(1)} | - \mu_1^2 \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{2\delta,j} \sigma_{2\delta,j+1} \right) \\ \stackrel{law}{=} \frac{1}{2\delta} \mu_1^{-4} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | \sigma_{\delta,2j} \epsilon_{\delta,2j} || \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} | - \mu_1^2 \sigma_{\delta,2j} \sigma_{\delta,2j+1} \right) \\ = \frac{1}{2\delta} \mu_1^{-4} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | \sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} || (\sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3}) | \right. \\ \left. - \mu_1^2 (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/2} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/2} \right) \end{aligned}$$

Assume that as $\delta \downarrow 0$ then $\sigma_{\delta,2j} = \sigma_{\delta,2j+1}$, therefore

$$\begin{aligned} \frac{1}{2\delta} \mu_1^{-4} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} | \sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1} || \sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3} | \right. \\ \left. - \mu_1^2 (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/2} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/2} \right) \\ \stackrel{law}{=} \frac{1}{2\delta} \mu_1^{-4} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{\delta,2j} \sigma_{\delta,2j+2} | \epsilon_{\delta,j} + \epsilon_{\delta,2j+1} || \epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3} | - 2\mu_1^2 \sigma_{\delta,2j} \sigma_{\delta,2j+2} \right) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{2\delta,j}^2 \sigma_{2\delta,j+1}^2 &= \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{2/2} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{2/2} \\ &\stackrel{law}{=} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} 4\sigma_{\delta,2j}^2 \sigma_{\delta,2j+2}^2 \\ &\xrightarrow{p} \int_0^t \sigma_s^2 ds. \end{aligned}$$

The variance can now be obtained

$$\begin{aligned}
\frac{1}{2\delta}\mu_1^{-4}Var & \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{\delta,2j}\sigma_{\delta,2j+2} \mid \epsilon_{\delta,j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2\sigma_{\delta,2j}\sigma_{\delta,2j+2} \right) \\
&= \frac{1}{2\delta}\mu_1^{-4} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{\delta,2j}^2\sigma_{\delta,2j+2}^2 \left(Var(\mid \epsilon_{\delta,j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2) \right. \\
&\quad \left. + 2Cov(\mid \epsilon_{\delta,j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2), \mid \epsilon_{\delta,j+2} + \epsilon_{\delta,2j+3} \mid \epsilon_{\delta,2j+4}\epsilon_{\delta,2j+5} \mid -2\mu_1^2) \right) \\
&= \frac{1}{2\delta}\mu_1^{-4}(C_1 + 2C_2) \sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{\delta,2j}^2\sigma_{\delta,2j+2}^2 \\
&\xrightarrow{p} 1/4\mu_1^{-4}(C_1 + 2C_2) \int_0^t \sigma_s^4 ds \\
&\simeq 2.592 \int_0^t \sigma_s^4 ds,
\end{aligned}$$

where

$$\begin{aligned}
C_1 &= Var(\mid u' + u'' \mid u''' + u^{(IV)} \mid) \\
C_2 &= Cov(\mid u' + u'' \mid u''' + u^{(IV)} \mid, \mid u''' + u^{(IV)} \mid \mid u^{(V)} + u^{(VI)} \mid).
\end{aligned}$$

The same procedure holds to obtain

$$\begin{aligned}
Var\left(\sqrt{\frac{1}{2\delta}}\left(\mu_1^{-2}[Y_{2\delta}^{(2)}]_t^{[1,1]} - \int_0^t \sigma_s^2 ds\right)\right) &\xrightarrow{p} 1/4\mu_1^{-4}(C_1 + 2C_2) \int_0^t \sigma_s^4 ds \\
&\simeq 2.592 \int_0^t \sigma_s^4 ds.
\end{aligned}$$

Now we should obtain the covariance. We need to assume that when $\delta \downarrow 0$ then $\sigma_{\delta,2j} = \sigma_{\delta,2j+1}$, so

$$\begin{aligned}
&Cov\left(\sqrt{\frac{1}{2\delta}}\left(\mu_1^{-2}[Y_{2\delta}^{(1)}]_t^{[1,1]} - \int_0^t \sigma_s^2 ds\right), \sqrt{\frac{1}{2\delta}}\left(\mu_1^{-2}[Y_{2\delta}^{(2)}]_t^{[1,1]} - \int_0^t \sigma_s^2 ds\right)\right) \\
&\stackrel{law}{=} \frac{1}{2\delta}\mu_1^{-4}Cov\left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 4} \sigma_{\delta,2j}\sigma_{\delta,2j+2}(\mid \epsilon_{\delta,2j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2), \right. \\
&\quad \left. \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} \sigma_{\delta,2j+1}\sigma_{\delta,2j+3} \mid \epsilon_{\delta,2j+1} + \epsilon_{\delta,2j+2} \mid \epsilon_{\delta,2j+3}\epsilon_{\delta,2j+4} \mid -2\mu_1^2)\right) \\
&\stackrel{law}{=} \frac{1}{2\delta}\mu_1^{-4} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} \sigma_{\delta,2j}^2\sigma_{\delta,2j+2}^2 \left(2Cov(\mid \epsilon_{\delta,2j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2, \right. \\
&\quad \mid \epsilon_{\delta,2j+1} + \epsilon_{\delta,2j+2} \mid \epsilon_{\delta,2j+3}\epsilon_{\delta,2j+4} \mid -2\mu_1^2) + 2Cov(\mid \epsilon_{\delta,2j} + \epsilon_{\delta,2j+1} \mid \epsilon_{\delta,2j+2}\epsilon_{\delta,2j+3} \mid -2\mu_1^2, \\
&\quad \mid \epsilon_{\delta,2j+3} + \epsilon_{\delta,2j+4} \mid \epsilon_{\delta,2j+5}\epsilon_{\delta,2j+6} \mid -2\mu_1^2) \Big) \\
&= \frac{1}{2\delta}\mu_1^{-4} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} \sigma_{\delta,2j}\sigma_{\delta,2j+2}(2C_3 + 2C_4)
\end{aligned}$$

$$\begin{aligned}
& \xrightarrow{p} 1/4\mu_1^{-4}(2C_3 + 2C_4) \int_0^t \sigma_s^4 ds \\
& \simeq 0.988 \int_0^t \sigma_s^4 ds
\end{aligned}$$

where

$$\begin{aligned}
C_3 &= Cov(|u' + u''| |u''' + u^{(IV)}|, |u'' + u'''| |u^{(IV)} + u^{(V)}|) \\
C_4 &= Cov(|u' + u''| |u''' + u^{(IV)}|, |u^{(IV)} + u^{(V)}| |u^{(VI)} + u^{(VII)}|).
\end{aligned}$$

Derivation for Realised Tripower Variation

We need to find the covariance matrix of

$$\sqrt{\frac{1}{2\delta}} \left(\mu_{2/3}^{-3} [Y_{2\delta}^{(1)}]_t^{[2/3, 2/3, 2/3]} - \int_0^t \sigma_s^2 ds \right)$$

If $\delta \downarrow 0$, we have that

$$\begin{aligned}
Var & \left(\sqrt{\frac{1}{2\delta}} \left(\mu_{2/3}^{-3} [Y_{2\delta}^{(1)}]_t^{[2/3, 2/3, 2/3]} - \int_0^t \sigma_s^2 ds \right) \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} |y_{2\delta, j}^{(1)}|^{2/3} |y_{2\delta, j+1}^{(1)}|^{2/3} |y_{2\delta, j+2}^{(1)}|^{2/3} - \mu_{2/3}^3 \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{2\delta, j}^{2/3} \sigma_{2\delta, j+1}^{2/3} \sigma_{2\delta, j+2}^{2/3} \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} |\sigma_{2\delta, j} \epsilon_{2\delta, j}|^{2/3} |\sigma_{2\delta, j+1} \epsilon_{2\delta, j+1}|^{2/3} |\sigma_{2\delta, j+2} \epsilon_{2\delta, j+2}|^{2/3} - \mu_{2/3}^3 \sigma_{2\delta, j}^{2/3} \sigma_{2\delta, j+1}^{2/3} \sigma_{2\delta, j+2}^{2/3} \right) \\
& = \frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} |\sigma_{\delta, 2j} \epsilon_{\delta, 2j} + \sigma_{\delta, 2j+1} \epsilon_{\delta, 2j+1}|^{2/3} (\sigma_{\delta, 2j+2} \epsilon_{\delta, 2j+2} + \sigma_{\delta, 2j+3} \epsilon_{\delta, 2j+3})^{2/3} \right. \\
& \quad \left. |(\sigma_{\delta, 2j+4} \epsilon_{\delta, 2j+4} + \sigma_{\delta, 2j+5} \epsilon_{\delta, 2j+5})|^{2/3} - \mu_{2/3}^3 (\sigma_{\delta, 2j}^2 + \sigma_{\delta, 2j+1}^2)^{2/6} (\sigma_{\delta, 2j+2}^2 + \sigma_{\delta, 2j+3}^2)^{2/6} \right. \\
& \quad \left. (\sigma_{\delta, 2j+4}^2 + \sigma_{\delta, 2j+5}^2)^{2/6} \right)
\end{aligned}$$

Assume that as $\delta \downarrow 0$ then $\sigma_{\delta, 2j} = \sigma_{\delta, 2j+1}$, so we obtain that

$$\begin{aligned}
& \frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} |\sigma_{\delta, 2j} \epsilon_{\delta, 2j} + \sigma_{\delta, 2j+1} \epsilon_{\delta, 2j+1}|^{2/3} |\sigma_{\delta, 2j+2} \epsilon_{\delta, 2j+2} + \sigma_{\delta, 2j+3} \epsilon_{\delta, 2j+3}|^{2/3} \right. \\
& \quad \left. |(\sigma_{\delta, 2j+4} \epsilon_{\delta, 2j+4} + \sigma_{\delta, 2j+5} \epsilon_{\delta, 2j+5})|^{2/3} - \mu_{2/3}^3 (\sigma_{\delta, 2j}^2 + \sigma_{\delta, 2j+1}^2)^{2/6} (\sigma_{\delta, 2j+2}^2 + \sigma_{\delta, 2j+3}^2)^{2/6} (\sigma_{\delta, 2j+4}^2 + \sigma_{\delta, 2j+5}^2)^{2/6} \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta, 2j}^{2/3} \sigma_{\delta, 2j+2}^{2/3} \sigma_{\delta, 2j+4}^{2/3} (|\epsilon_{\delta, j} + \epsilon_{\delta, 2j+1}|^{2/3} |\epsilon_{\delta, 2j+2} \epsilon_{\delta, 2j+3}|^{2/3} |\epsilon_{\delta, j+4} + \epsilon_{\delta, 2j+5}|^{2/3} \right. \\
& \quad \left. - 2\mu_{2/3}^3) \right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{2\delta,j}^{4/3} \sigma_{2\delta,j+1}^{4/3} \sigma_{2\delta,j+2}^{4/3} &= \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{2/3} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{2/3} (\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)^{2/3} \\
&\stackrel{law}{=} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} 4\sigma_{\delta,2j}^{4/3} \sigma_{\delta,2j+2}^{4/3} \sigma_{\delta,2j+4}^{4/3} \\
&\xrightarrow{p} \int_0^t \sigma_s^4 ds.
\end{aligned}$$

So then we find that

$$\begin{aligned}
&\frac{1}{2\delta} \mu_{2/3}^{-6} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta,2j}^{2/3} \sigma_{\delta,2j+2}^{2/3} \sigma_{\delta,2j+4}^{2/3} (|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} \right. \\
&\quad \left. - 2\mu_{2/3}^3) \right) \\
&= \frac{1}{2\delta} \mu_{2/3}^{-6} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta,2j}^{4/3} \sigma_{\delta,2j+2}^{4/3} \sigma_{\delta,2j+4}^{4/3} \\
&\quad \times \left(Var(|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3) \right. \\
&\quad + 2Cov(|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3, \\
&\quad \quad |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} |\epsilon_{\delta,2j+6} + \epsilon_{\delta,2j+7}|^{2/3} - 2\mu_{2/3}^3) \\
&\quad + 2Cov(|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3, \\
&\quad \quad |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} |\epsilon_{\delta,2j+6} + \epsilon_{\delta,2j+7}|^{2/3} |\epsilon_{\delta,2j+8} + \epsilon_{\delta,2j+9}|^{2/3} - 2\mu_{2/3}^3) \Big) \\
&= \frac{1}{2\delta} \mu_{2/3}^{-6} (C_5 + 2C_6 + 2C_7) \sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta,2j}^{4/3} \sigma_{\delta,2j+2}^{4/3} \sigma_{\delta,2j+4}^{4/3} \\
&\xrightarrow{p} 1/4 \mu_{2/3}^{-6} (C_5 + 2C_6 + 2C_7) \int_0^t \sigma_s^4 ds \\
&\simeq 3.049 \int_0^t \sigma_s^4 ds,
\end{aligned}$$

where

$$\begin{aligned}
C_5 &= Var(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}) \\
C_6 &= Cov(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}, \\
&\quad |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3} |u^{(VII)} + u^{(VIII)}|^{2/3}), \\
C_7 &= Cov(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}, \\
&\quad |u^{(V)} + u^{(VI)}|^{2/3} |u^{(VII)} + u^{(VIII)}|^{2/3} |u^{(IX)} + u^{(X)}|^{2/3}).
\end{aligned}$$

The same procedure holds to obtain

$$Var \left(\sqrt{\frac{1}{2\delta}} \left(\mu_{2/3}^{-6} [Y_{2\delta}^{(2)}]_t^{[2/3, 2/3, 2/3]} - \int_0^t \sigma_s^2 ds \right) \right) \xrightarrow{p} 1/4 \mu_{2/3}^{-6} (C_5 + 2C_6 + 2C_7) \int_0^t \sigma_s^4 ds$$

$$\simeq 3.049 \int_0^t \sigma_s^4 ds.$$

Now to obtain the covariance, we need to assume that when $\delta \downarrow 0$ then $\sigma_{\delta,2j} = \sigma_{\delta,2j+1}$, so

$$\begin{aligned} & Cov\left(\sqrt{\frac{1}{2\delta}}\left(\mu_{2/3}^{-3}[Y_{2\delta}^{(1)}]_t^{[2/3,2/3,2/3]} - \int_0^t \sigma_s^2 ds\right), \sqrt{\frac{1}{2\delta}}\left(\mu_{2/3}^{-3}[Y_{2\delta}^{(2)}]_t^{[2/3,2/3,2/3]} - \int_0^t \sigma_s^2 ds\right)\right) \\ & \stackrel{law}{=} \frac{1}{2\delta} \mu_{2/3}^{-6} Cov\left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta,2j}^{2/3} \sigma_{\delta,2j+2}^{2/3} \sigma_{\delta,2j+4}^{2/3} (|\epsilon_{\delta,2j} + \epsilon_{\delta,2j+1}|^{\frac{2}{3}} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{\frac{2}{3}} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{\frac{2}{3}} - 2\mu_{2/3}^3), \right. \\ & \quad \left. \sum_{j=0}^{\lfloor t/2\delta \rfloor - 7} \sigma_{\delta,2j+1}^{2/3} \sigma_{\delta,2j+3}^{2/3} \sigma_{\delta,2j+5}^{2/3} (|\epsilon_{\delta,2j+1} + \epsilon_{\delta,2j+2}|^{2/3} |\epsilon_{\delta,2j+3} + \epsilon_{\delta,2j+4}|^{2/3} |\epsilon_{\delta,2j+5} + \epsilon_{\delta,2j+6}|^{2/3} - 2\mu_{2/3}^3) \right) \\ & \stackrel{law}{=} \frac{1}{2\delta} \mu_{2/3}^{-6} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 7} \sigma_{\delta,2j}^{4/3} \sigma_{\delta,2j+2}^{4/3} \sigma_{\delta,2j+4}^{4/3} \left(2Cov(|\epsilon_{\delta,2j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3, \right. \\ & \quad |\epsilon_{\delta,2j+1} + \epsilon_{\delta,2j+2}|^{2/3} |\epsilon_{\delta,2j+3} + \epsilon_{\delta,2j+4}|^{2/3} |\epsilon_{\delta,2j+5} + \epsilon_{\delta,2j+6}|^{2/3} - 2\mu_{2/3}^3) \\ & \quad + 2Cov(|\epsilon_{\delta,2j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3, \\ & \quad |\epsilon_{\delta,2j+3} + \epsilon_{\delta,2j+4}|^{2/3} |\epsilon_{\delta,2j+5} + \epsilon_{\delta,2j+6}|^{2/3} |\epsilon_{\delta,2j+7} + \epsilon_{\delta,2j+8}|^{2/3} - 2\mu_{2/3}^3) \\ & \quad + 2Cov(|\epsilon_{\delta,2j} + \epsilon_{\delta,2j+1}|^{2/3} |\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}|^{2/3} |\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}|^{2/3} - 2\mu_{2/3}^3, \\ & \quad \left. |\epsilon_{\delta,2j+5} + \epsilon_{\delta,2j+6}|^{2/3} |\epsilon_{\delta,2j+7} + \epsilon_{\delta,2j+8}|^{2/3} |\epsilon_{\delta,2j+9} + \epsilon_{\delta,2j+10}|^{2/3} - 2\mu_{2/3}^3) \right) \\ & = \frac{1}{2\delta} \mu_{2/3}^{-6} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 5} \sigma_{\delta,2j}^{4/3} \sigma_{\delta,2j+2}^{4/3} \sigma_{\delta,2j+4}^{4/3} (2C_8 + 2C_9 + 2C_{10}) \\ & \xrightarrow{p} 1/4 \mu_{2/3}^{-6} (2C_8 + 2C_9 + 2C_{10}) \int_0^t \sigma_s^4 ds \\ & \simeq 1.028 \int_0^t \sigma_s^4 ds \end{aligned}$$

where

$$\begin{aligned} C_8 &= Cov(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}, \\ & \quad |u'' + u'''|^{2/3} |u^{(IV)} + u^{(V)}|^{2/3} |u^{(VI)} + u^{(VII)}|^{2/3}) \\ C_9 &= Cov(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}, \\ & \quad |u^{(IV)} + u^{(V)}|^{2/3} |u^{(VI)} + u^{(VII)}|^{2/3} |u^{(VII)} + u^{(IX)}|^{2/3}) \\ C_{10} &= Cov(|u' + u''|^{2/3} |u''' + u^{(IV)}|^{2/3} |u^{(V)} + u^{(VI)}|^{2/3}, \\ & \quad |u^{(VI)} + u^{(VII)}|^{2/3} |u^{(VIII)} + u^{(IX)}|^{2/3} |u^{(X)} + u^{(XI)}|^{2/3}). \end{aligned}$$

Derivation for Realised Quadpower Variation

We need to find the covariance matrix of

$$\sqrt{\frac{1}{2\delta}} \begin{pmatrix} \mu_{1/2}^{-4} [Y_{2\delta}^{(1)}]_t^{[1/2,1/2,1/2,1/2]} - \int_0^t \sigma_s^2 ds \\ \mu_{1/2}^{-4} [Y_{2\delta}^{(2)}]_t^{[1/2,1/2,1/2,1/2]} - \int_0^t \sigma_s^2 ds \end{pmatrix}.$$

If $\delta \downarrow 0$, we have that

$$\begin{aligned}
Var & \left(\sqrt{\frac{1}{2\delta}} \left(\mu_{1/2}^{-4} [Y_{2\delta}^{(1)}]_t^{[1/2, 1/2, 1/2, 1/2]} - \int_0^t \sigma_s^2 ds \right) \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{1/2}^{-8} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sqrt{|y_{2\delta,j}^{(1)}| |y_{2\delta,j+1}^{(1)}| |y_{2\delta,j+2}^{(1)}| |y_{2\delta,j+3}^{(1)}|} - \mu_{1/2}^4 \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sqrt{\sigma_{2\delta,j} \sigma_{2\delta,j+1} \sigma_{2\delta,j+2} \sigma_{2\delta,j+3}} \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{1/2}^{-8} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} |\sigma_{2\delta,j} \epsilon_{2\delta,j}|^{1/2} |\sigma_{2\delta,j+1} \epsilon_{2\delta,j+1}|^{1/2} |\sigma_{2\delta,j+2} \epsilon_{2\delta,j+2}|^{1/2} |\sigma_{2\delta,j+1} \epsilon_{2\delta,j+1}|^{1/2} \right. \\
& \quad \left. - \mu_{1/2}^4 \sigma_{2\delta,j}^{1/2} \sigma_{2\delta,j+1}^{1/2} \sigma_{2\delta,j+2}^{1/2} \sigma_{2\delta,j+3}^{1/2} \right) \\
& = \frac{1}{2\delta} \mu_{1/2}^{-8} Var \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} |\sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1}|^{1/2} |(\sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3})|^{1/2} \right. \\
& \quad |(\sigma_{\delta,2j+4} \epsilon_{\delta,2j+4} + \sigma_{\delta,2j+5} \epsilon_{\delta,2j+5})|^{1/2} |(\sigma_{\delta,2j+6} \epsilon_{\delta,2j+6} + \sigma_{\delta,2j+7} \epsilon_{\delta,2j+7})|^{1/2} \\
& \quad \left. - \mu_{1/2}^4 (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/4} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/4} (\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)^{1/4} (\sigma_{\delta,2j+6}^2 + \sigma_{\delta,2j+7}^2)^{1/4} \right)
\end{aligned}$$

Assume that as $\delta \downarrow 0$ then $\sigma_{\delta,2j} = \sigma_{\delta,2j+1}$, so

$$\begin{aligned}
\frac{1}{2\delta} \mu_{1/2}^{-8} Var & \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} |\sigma_{\delta,2j} \epsilon_{\delta,2j} + \sigma_{\delta,2j+1} \epsilon_{\delta,2j+1}|^{1/2} |\sigma_{\delta,2j+2} \epsilon_{\delta,2j+2} + \sigma_{\delta,2j+3} \epsilon_{\delta,2j+3}|^{1/2} \right. \\
& \quad |(\sigma_{\delta,2j+4} \epsilon_{\delta,2j+4} + \sigma_{\delta,2j+5} \epsilon_{\delta,2j+5})|^{1/2} |(\sigma_{\delta,2j+6} \epsilon_{\delta,2j+6} + \sigma_{\delta,2j+7} \epsilon_{\delta,2j+7})|^{1/2} \\
& \quad \left. - \mu_{1/2}^4 (\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)^{1/4} (\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)^{1/4} (\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)^{1/4} (\sigma_{\delta,2j+6}^2 + \sigma_{\delta,2j+7}^2)^{1/4} \right) \\
& \stackrel{law}{=} \frac{1}{2\delta} \mu_{1/2}^{-8} Var \left(\sum_{i=0}^{\lfloor t/2\delta \rfloor - 6} \sigma_{\delta,2j}^{1/2} \sigma_{\delta,2j+2}^{1/2} \sigma_{\delta,2j+4}^{1/2} \sigma_{\delta,2j+6}^{1/2} (|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{1/2} |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}|^{1/2} \right. \\
& \quad \left. |\epsilon_{\delta,j+4} + \epsilon_{\delta,2j+5}|^{1/2} |\epsilon_{\delta,j+5} + \epsilon_{\delta,2j+7}|^{1/2} - 2\mu_{1/2}^4) \right)
\end{aligned}$$

therefore

$$\begin{aligned}
& \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sigma_{2\delta,j} \sigma_{2\delta,j+1} \sigma_{2\delta,j+2} \sigma_{2\delta,j+3} \\
& = \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sqrt{(\sigma_{\delta,2j}^2 + \sigma_{\delta,2j+1}^2)(\sigma_{\delta,2j+2}^2 + \sigma_{\delta,2j+3}^2)(\sigma_{\delta,2j+4}^2 + \sigma_{\delta,2j+5}^2)(\sigma_{\delta,2j+6}^2 + \sigma_{\delta,2j+7}^2)} \\
& \stackrel{law}{=} \frac{1}{2\delta} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} 4\sigma_{\delta,2j} \sigma_{\delta,2j+2} \sigma_{\delta,2j+4} \sigma_{\delta,2j+6} \\
& \xrightarrow{p} \int_0^t \sigma_s^4 ds.
\end{aligned}$$

So now the variance can be obtained

$$\begin{aligned}
\frac{1}{2\delta}\mu_{1/2}^{-8}Var & \left(\sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sigma_{\delta,2j}^{1/2} \sigma_{\delta,2j+2}^{1/2} \sigma_{\delta,2j+4}^{1/2} \sigma_{\delta,2j+6}^{1/2} (|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}|^{1/2} |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}|^{1/2} \right. \\
& \quad \left. |\epsilon_{\delta,2j+3} \epsilon_{\delta,2j+4}|^{1/2} |\epsilon_{\delta,2j+5} \epsilon_{\delta,2j+6}|^{1/2} - 2\mu_{1/2}^4) \right) \\
&= \frac{1}{2\delta}\mu_{1/2}^{-8} \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sigma_{\delta,2j} \sigma_{\delta,2j+2} \sigma_{\delta,2j+4} \sigma_{\delta,2j+6} \left(\right. \\
& \quad Var(\sqrt{|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}| |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}| |\epsilon_{\delta,2j+4} \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}|} - 2\mu_{1/2}^4) \\
& \quad + 2Cov(\sqrt{|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}| |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}| |\epsilon_{\delta,2j+4} \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}|} - 2\mu_{1/2}^4), \\
& \quad \sqrt{|\epsilon_{\delta,2j+2} + \epsilon_{\delta,2j+3}| |\epsilon_{\delta,2j+4} \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}| |\epsilon_{\delta,2j+8} \epsilon_{\delta,2j+9}|} - 2\mu_{1/2}^4) \\
& \quad + 2Cov(\sqrt{|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}| |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}| |\epsilon_{\delta,2j+4} \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}|} - 2\mu_{1/2}^4), \\
& \quad \sqrt{|\epsilon_{\delta,2j+4} + \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}| |\epsilon_{\delta,2j+8} \epsilon_{\delta,2j+9}| |\epsilon_{\delta,2j+10} \epsilon_{\delta,2j+11}|} - 2\mu_{1/2}^4) \\
& \quad + 2Cov(\sqrt{|\epsilon_{\delta,j} + \epsilon_{\delta,2j+1}| |\epsilon_{\delta,2j+2} \epsilon_{\delta,2j+3}| |\epsilon_{\delta,2j+4} \epsilon_{\delta,2j+5}| |\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}|} - 2\mu_{1/2}^4), \\
& \quad \left. \sqrt{|\epsilon_{\delta,2j+6} \epsilon_{\delta,2j+7}| |\epsilon_{\delta,2j+8} \epsilon_{\delta,2j+9}| |\epsilon_{\delta,2j+10} \epsilon_{\delta,2j+11}| |\epsilon_{\delta,2j+12} + \epsilon_{\delta,2j+13}|} - 2\mu_{1/2}^4) \right) \\
&= \frac{1}{2\delta}\mu_{1/2}^{-8} (C_{11} + 2C_{12} + 2C_{13} + 2C_{14}) \sum_{j=0}^{\lfloor t/2\delta \rfloor - 8} \sigma_{\delta,2j} \sigma_{\delta,2j+2} \sigma_{\delta,2j+4} \sigma_{\delta,2j+6} \\
&\xrightarrow{p} 1/4\mu_{1/2}^{-8} (C_{11} + 2C_{12} + 2C_{13} + 2C_{14}) \int_0^t \sigma_s^4 ds \\
&\simeq 3.378 \int_0^t \sigma_s^4 ds,
\end{aligned}$$

where

$$\begin{aligned}
C_{11} &= Var(\sqrt{|u' + u''| |u''' + u^{(IV)}| |u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}|}) \\
C_{12} &= Cov(\sqrt{|u' + u''| |u''' + u^{(IV)}| |u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}|}, \\
& \quad \sqrt{|u''' + u^{(IV)}| |u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}| |u^{(IX)} + u^{(X)}|}). \\
C_{13} &= Cov(\sqrt{|u' + u''| |u''' + u^{(IV)}| |u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}|}, \\
& \quad \sqrt{|u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}| |u^{(IX)} + u^{(X)}| |u^{(XI)} + u^{(XII)}|}). \\
C_{14} &= Cov(\sqrt{|u' + u''| |u''' + u^{(IV)}| |u^{(V)} + u^{(VI)}| |u^{(VII)} + u^{(VIII)}|}, \\
& \quad \sqrt{|u^{(VII)} + u^{(VIII)}| |u^{(IX)} + u^{(X)}| |u^{(XI)} + u^{(XII)}| |u^{(XIII)} + u^{(XIV)}|}).
\end{aligned}$$

The same procedure holds to obtain

$$\begin{aligned}
Var\left(\sqrt{\frac{1}{2\delta}}\left(\mu_{1/2}^{-8}[Y_{2\delta}^{(2)}]_t^{[1/2,1/2,1/2,1/2]} - \int_0^t \sigma_s^2 ds\right)\right) &\xrightarrow{p} 1/4\mu_{1/2}^{-8}(C_{11} + 2C_{12} + 2C_{13} + 2C_{14}) \int_0^t \sigma_s^4 ds \\
&\simeq 3.378 \int_0^t \sigma_s^4 ds.
\end{aligned}$$

Now to obtain the covariance, we need to assume that when $\delta \downarrow 0$ then $\sigma_{\delta,2j} = \sigma_{\delta,2j+1}$, so

References

- Aït-Sahalia Y., P. Mykland and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18, 351-416.
- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economics Review* 39, 885-905.
- Andersen, T. G., T. Bollerslev and F. X. Diebold (2005). Parametric and nonparametric measurements of volatility. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: Elsevier Science B.V.
- Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (1999). Microstructure bias and volatility signatures. Unpublished paper.
- Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2000). Great Realizations. *Risk* 13, 105-108.
- Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2001). The distribution of exchange rates volatility. *Journal of American Statistical Association* 96, 42-55.
- Bai, X., J. R. Russell and G. C. Tiao (2001). Beyond Merton's Utopia (1): effects of non-normality and dependence on the precision of variance estimates using high frequency data. Unpublished paper: University of Chicago.
- Bandi, F. M. and J. R. Russell (2003). Microstructures noise, realized volatility and optimal sampling. Unpublished paper presented at the Realized volatility conference, Montreal, 8th November 2003.
- Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij and N. Shephard (2005). A central limit theorem for realised power and bipower variations of continuous semimartingales. In Y. Kabanov and R. Lipster (Eds.), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*. New York: Springer Verlag.
- Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod and N. Shephard (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22, 677-719.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde and N. Shephard (2006). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Unpublished paper: Nuffield College, Oxford.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B* 63, 167-241.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253-280.
- Barndorff-Nielsen, O. E. and N. Shephard (2003). Realised power variation and stochastic volatility. *Bernoulli* 9, 243-265.
- Barndorff-Nielsen O. E. and N. Shephard (2004a). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2, 1-48.
- Barndorff-Nielsen, O. E. and N. Shephard (2004b). Econometric analysis of realised covariation:

high frequency covariance, regression and correlation in financial economics. *Econometrica* 72, 885-925.

Barndorff-Nielsen O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1-30.

Barndorff-Nielsen O. E. and N. Shephard (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. In R. Blundell, P. Torsten and W. K. Newey (Eds.), *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress. Econometric Society Monographs*. Cambridge: Cambridge University Press.

Christensen, B. J. and N. R. Prabhala (1998). The relation between implied and realized volatility. *Journal of Financial Economics* 37, 125-150.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8, 291-323.

Corsi, F., G. Zumbach, U. Müller and M. Dacorogna (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes* 30, 183-204.

Cox, J. C., J. E. Ingersoll and S. A. Ross (1985). A theory of the term structures of interest rates. *Econometrica* 53, 385-407.

Dacorogna, M. M., R. Gencay, U. A. Müller, R. B. Olsen and O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics* 22, no. 3, 427-432.

Hansen, P. R. and A. Lunde (2004). An unbiased measure of realized variance. Brown University Working Paper.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*. Forthcoming.

Hsieh, D. (1991). Chaos and nonlinear dynamics: Applications to financial markets. *Journal of Finance* 42, 281-300.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance* 42, 281-300.

Lo, A. W. and A. C. MacKinlay (1990). An econometric analysis of non-synchronous trading. *Journal of Econometrics* 45, 181-212.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323-361.

Oomen, R. C. A. (2002). Modelling realized variance when returns are serially correlated. Unpublished paper: Warwick Business School, University of Warwick.

Oomen, R. C. A. (2004). Properties of bias corrected realized variance in calendar time and business time. Unpublished paper: Warwick Business School, University of Warwick.

Poterba, J. and L. Summers (1986). The persistence of volatility and stock market fluctuations. *American Economic Review* 76, 1124-1141.

Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance* 39, 1127-1139.

Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security prices. Working paper 11, Graduate School of Business Administration, University

of California, Berkely. Reprinted in Shephard (2005).

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1115-1153.

Taylor, S. J. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317-340.

Ysusi, C. (2006). Detecting jumps in high-frequency financial series using multipower variation. Working paper 2006-10, Banco de México.

Zhang, L., P. Mykland and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of American Statistical Association* 100, 1394-1411.

Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics* 14, 45-52.